



ANAIS DA JORNADA CIENTÍFICA E DE EXTENSÃO (JCE 2017)

**Coordenação Geral da JCE
Patricia Endo e Wylliams Santos**

Analisando o impacto das falhas da infraestrutura de energia na disponibilidade de aplicativos em nuvem

Élisson Rocha¹, Patricia Takako Endo¹

¹Grupo de Estudos Avançados em Tecnologia da Informação e Comunicação (GREAT)
Universidade Pernambuco (UPE) - Campus Caruaru
BR 104, S/N – Caruaru – PE – Brasil

elissonrochaa@gmail.com, patricia.endo@upe.br

Resumo. *A adoção generalizada da computação em nuvem tem sido impulsionada por vantagens percebidas em agilidade e escalabilidade inerentes à computação em nuvem. Para os provedores de serviço em nuvem, gerenciar a alta expectativa dos clientes e os danos resultantes das interrupções do serviço é um desafio crescente e crítico. Este trabalho examina a relação entre arquiteturas de infraestrutura de energia e disponibilidade de aplicativos de software em um data center. Os resultados apontam que a infraestrutura de energia é importante na disponibilidade do serviço. À medida que aumentamos o nível de redundância da arquitetura, conseguimos melhores resultados.*

1. Introdução

Com o intuito de se ampliar a abrangência das aplicações para Cidades Criativas e considerando a limitação dos recursos computacionais dos dispositivos móveis, pode-se transferir parte do processamento destas aplicações (ou até mesmo toda a aplicação) para o ambiente de Nuvem Computacional. De acordo com o NIST, Computação em Nuvem pode ser definido como “um modelo computacional que permite de maneira onipresente e conveniente o acesso a dados e informações disponíveis na rede”. Através deste modelo computacional, é possível compartilhar recursos computacionais de forma rápida, podendo ser provisionados e liberados com o mínimo de esforço de gerenciamento e interação com os provedores dos serviços prestados. Baseado nessas características, pode-se identificar como ponto principal dos ambientes em nuvem, o compartilhamento e a integração de recursos distribuídos para o aumento e disponibilidade da capacidade de processamento de um determinado serviço [Endo et al. 2011].

Como resultado dessa alta demanda, os provedores especializados, atualmente baseado em grandes data centers partem para estratégias com grandes níveis de tolerância a falhas para atender as expectativas cada vez mais altas dos clientes e seus usuários. No entanto, um princípio utilizado na tolerância a falhas é introduzir redundância, que nada mais é que o uso de componentes, que individualmente tem disponibilidade baixa, em agregado para fornecer alta disponibilidade. Esse método trás consigo um número maior de dispositivos para a infraestrutura, conseqüentemente, mais recursos de energia será necessário, o que pode ser um ponto negativo para a sustentabilidade e para o custo de aquisição e manutenção do sistema. Desta forma, a confiabilidade, disponibilidade, sustentabilidade e custos do sistemas de data center são preocupações reais da comunidade acadêmica e sociedade em geral.

A disponibilidade do serviço pode ser afetada por vários fatores, tais como falhas de componentes de TI, ou em infraestruturas de arrefecimento e de energia; E pode piorar devido à complexidade das conexões desses componentes. Entre todos os subsistemas que compõem um centro de dados, a infraestrutura de energia é crucial para manter todos os serviços e outros subsistemas operando em condições operacionais. Tipicamente, uma infraestrutura de energia é composta por uma subestação de serviço público, um sistema de fonte de alimentação ininterrupta (UPS), geradores e unidades de distribuição de energia (PDUs); E qualquer falha nestes componentes pode levar à indisponibilidade do data center como um todo.

Foi proposto um conjunto de arquiteturas da infraestrutura de energia com base na análise de sensibilidade, e representando as mesmas utilizando modelos de Rede de Petri estocástica (*Stochastic Petri Net - SPN*) [Reisig 2013]. O principal objetivo é analisar os impactos de falhas da infraestrutura de energia na disponibilidade dos aplicativos hospedados em um data center. Para isso, também foi proposto um modelo genérico para representar as falhas dos equipamentos de TI usando o Diagrama de Blocos de Confiabilidade (*Reliability Block Diagram - RBD*) [Cepin 2011] e SPN para falhas na aplicação com replicação baseada em *checkpoints*.

2. Modelos de disponibilidade

Esta seção está dividida em três modelos distintos: infraestrutura de energia (subseção 2.1), aplicação com mecanismo de checkpoint (subseção 2.2) e Infraestrutura de TI com a integração dos três modelos (subseção 2.3; e na 2.4 é mostrado os resultados dessa integração.

2.1. Infraestrutura de energia

A infraestrutura de energia é composta por muitos componentes: AC Source (AC), Generator (G), Automatic Transfer Switch (ATS), Voltage Panel (VP), Uninterruptible Power Supply (UPS), Transformer (T), SubPanel (SP) e Junction Box (JB), e podem ser organizados em série. Figura 1 apresenta nossa arquitetura básica de energia (chamada A0), sem redundância. A infraestrutura de energia falha quando um componente do sistema está danificado, exceto para o AC e G. Por esse motivo, a arquitetura A0 é vulnerável a qualquer falha.

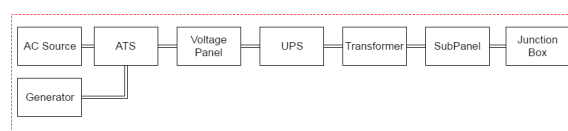


Figura 1. Arquitetura A0 - infraestrutura básica de energia sem redundância

O modelo SPN correspondente a A0 é mostrado na Figura 2. Cada equipamento é representado por um bloco de construção com dois lugares e duas transições, componente simples. Os lugares denominados *componente_UP* e *componente_DOWN* representam componentes funcionando ou em falha, respectivamente. A transição que representa uma falha tem um valor MTTF, enquanto a transição que representa um reparo possui valores MTTR.

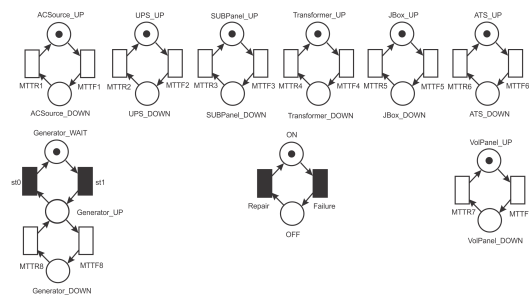


Figura 2. SPN da Arquitetura A0

O componente Gerador varia em apenas um respeito: além do componente simples mencionado anteriormente, apresenta mais um lugar, chamado *WAIT* e duas transições imediatas (*st0* e *st1*). O gerador só funciona quando o AC Source falha; Logo, se isso não acontecer, ele fica aguardando, no estado *WAIT*. Quando o AC Source falhar, o gerador é ativado (transição *st1*), e quando a fonte de CA retorna, o gerador retorna ao status de espera (transição *st0*).

As funções de guarda presentes nas transições imediatas são descritas na Tabela 1. Além dos componentes da arquitetura de energia, o estado do sistema geral é representado com outro bloco de construção, com dois lugares, *ON* e *OFF* e duas transições imediatas com funções de guarda, *Failure* E *Repair*, que é visto suas funções de guarda mais a frente.

Tabela 1. Funções de guarda de transição imediata do SPN - A0

Identificação	Função de Guarda
<i>st0</i>	#ACSource_DOWN=0
<i>st1</i>	#ACSource_DOWN=1

Para melhorar esta arquitetura, foi realizado uma análise de sensibilidade para entender qual componente tem um grande impacto na disponibilidade geral do sistema. Os valores MTTF e MTTR de cada componente utilizados na análise de sensibilidade estão descritos na Tabela 2.

Tabela 2. Valores MTTF e MTTR dos componentes do sistema de energia (por [Callou et al. 2013])

Componente	MTTF (em hora)	MTTR (em hora)
AC Source	4380	8
Generator	2190	8
UPS	50000	8
SubPanel	304000	8
Transformer	282581	8
Junction Box	5244000	8
Voltage panel	304000	8
ATS	48076	8

Tabela 3 mostra os resultados da análise de sensibilidade de A0, com os cinco componentes que têm mais impacto na disponibilidade. Como se pode notar, o ATS apresenta o maior índice (0.0000383942371), o que significa que ele tem mais impacto na disponibilidade da infraestrutura de energia.

Tabela 3. Análise de sensibilidade da arquitetura A0

Componente	Índice de sensibilidade
ATS	0.0000383942371
UPS	0.0000369175256
Voltage Panel	0.0000060728889
SubPanel	0.0000060728813
AC Source	0.00000076745982

Então, considerando o resultado da análise de sensibilidade, propomos a arquitetura A1 com dois componentes ATS, como mostrado na Figura 3.

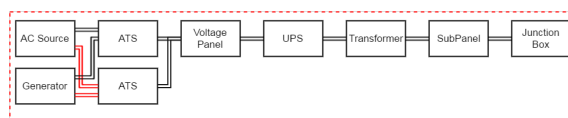


Figura 3. Arquitetura A1 - Infraestrutura de energia com redundância no ATS

O modelo SPN correspondente à arquitetura A1 é semelhante ao A0, como mostrado na Figura 4. A diferença é que temos dois tokens no lugar *ATS_UP* que representam a redundância ATS no modelo e as políticas de concorrência, ao invés de ser serviço único, é de serviço infinito.

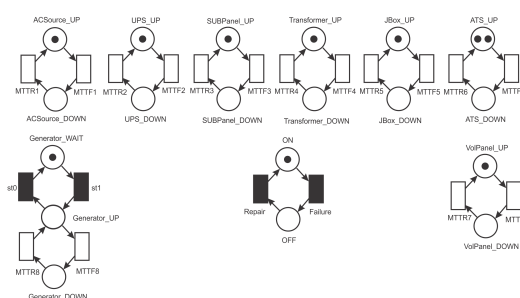


Figura 4. SPN da Arquitetura A1

Para propor novas arquiteturas de energia, seguimos a metodologia baseada na análise de sensibilidade, replicando o componente que atingiu o maior índice, conforme descrito para propor a arquitetura A1. Desta forma, temos quatro modelos de arquitetura de energia, e os componentes de energia de cada modelo são descritos na Tabela 4.

Os modelos SPN de A2 e A3 são semelhantes a A1, onde para cada componente de redundância, temos dois tokens para representá-lo. Na arquitetura A2, temos ATS e UPS com redundância; E na A3, temos ATS, UPS, VP, T e SP.

Tabela 4. Arquiteturas de energia e seus componentes

Arquitetura	Componentes
A0	AC-G-ATS-VP-UPS-T-SP-JB
A1	AC-G-2ATS-VP-UPS-T-SP-JB
A2	AC-G-2ATS-VP-2UPS-T-SP-JB
A3	AC-G-2ATS-2VP-2UPS-2T-2SP-JB

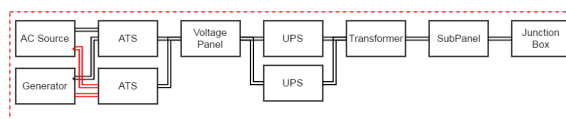


Figura 5. Arquitetura A2 - infraestrutura de energia com redundância ATS e UPS

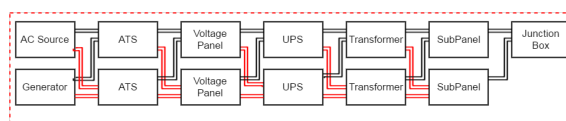


Figura 6. Arquitetura A3 - Infraestrutura de energia com redundância ATS, VP, UPS, T e SP

Na Arquitetura A3, a redundância é em quase todos os componentes, já que o gerador serve como redundância para o AC Source. Com isso, só o Junction Box fica sem redundância, pois, como visto na tabela 2, o seu MTTF é muito alto, com isso faz com que o componente tenha um pequeno impacto na disponibilidade.

2.2. Aplicação com modelo de checkpoint

O checkpoint é um mecanismo que auxilia na busca da alta disponibilidade, é o processo de salvar as informações necessárias do estado atual para manter o aplicativo em execução mais tarde [Singh et al. 2012]. Com isso, em caso de falha, a aplicação pode ser reiniciada com perda mínima de dados a partir do último *checkpoint* salvo. Na Figura 7 é visto o modelo SPN do checkpoint.

A aplicação tem dois estados, UP e DOWN, possui também uma transição MTTF10, onde nela consta o tempo médio de falha do componente, e ao invés de considerar o MTTR da aplicação, utiliza-se os valores de recuperação do mecanismo de checkpoint.

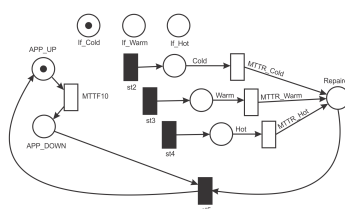


Figura 7. Modelo de Checkpoint no SPN

Quando a aplicação falha, estado *APP_DOWN*, começa o mecanismo de checkpoint. Primeiro, com os places *If_Cold*, *If_Warm* e *If_Hot*, é possível saber qual

modo de recuperação vai ser utilizado, isso porque para cada modo de recuperação é um valor de MTTR diferente, *MTTR_Cold*, *MTTR_Warm* e *MTTR_Hot*. Definimos funções de guarda para garantir esse comportamento, como se pode ver na tabela 5.

Tabela 5. Função de Guarda do modelo de checkpoint

Indentificação	Função de Guarda
st2	(#App_UP=0)AND(#If_Cold=1)AND(#Cold=0)AND(#Repaired=0)
st3	(#App_UP=0)AND(#If_Warm=1)AND(#Warm=0)AND(Repaired=0)
st4	(#App_UP=0)AND(#If_Hot=1)AND(#Hot=0)AND(#Repaired=0)

Após definido o modo de recuperação, o token vai para o place correspondente ao modo selecionado e habilita a transição MTTR. Passado o tempo médio, a aplicação é reparada e voltada para o estado *APP_UP*.

2.3. Integração com o modelo da infraestrutura de energia

Com o mecanismo de checkpoint anteriormente apresentado se refere a aplicação, e tem-se também a infraestrutura de TI. A infraestrutura de TI será utilizado o RBD, como pode ser visto na Figura 8.



Figura 8. Infraestrutura de TI no RBD

A infraestrutura de TI será formado por Hardware, Sistema Operacional e Máquina Virtual. Os valores dos componentes são os citados na Tabela 6, é visto também os valores para as aplicações Video por Demanda e Livraria Digital. Com o valor de MTTF e MTTR gerado por esse RBD, ver Tabela 7, será criado, no SPN, um componente simples. Esse componente representará a infraestrutura de TI na integração com a infraestrutura de energia e a aplicação.

Tabela 6. MTTF e MTTR dos componentes de TI. Por [Melo et al. 2015, Araujo et al. 2014]

Componente	MTTF (em hora)	MTTR (em hora)
Hardware	8760	1.667
Sistema Operacional	1440	1
Máquina virtual	1880	0.167
Aplicação de video por demanda	217.77	0.92633
Aplicação de Livraria Digital	6865.3	0.167

Tabela 7. Valores de MTTF e MTTR da InfraTI

Valores	
MTTF	745.98218
MTTR	0.726424

Com o componente simples adicionado e o modelo de checkpoint, é possível ver na figura 9 a integração do modelo da infraestrutura de energia A0, com a infraestrutura de TI e aplicação com checkpoint (a integração é a mesma para próximas arquiteturas de energia).

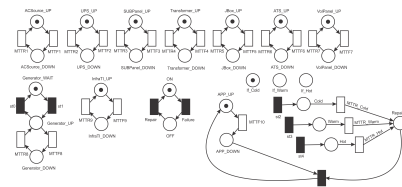


Figura 9. SPN da integração

Na integração é preciso atualizar as funções de guarda das transições imediatas Failure e Repair. Pode ser visto as funções atualizadas na Tabela 8.

Tabela 8. Funções de guarda do sistema geral

Identificação	Função de guarda
<i>Failure</i>	$((\#ACSource_UP=0) \text{ AND } (\#Generator_UP=0)) \text{ OR } (\#ATS_UP=0) \text{ OR } (\#VoltagePanel_UP=0) \text{ OR } (\#UPS_UP=0) \text{ OR } (\#Transformer_UP=0) \text{ OR } (\#SubPanel_UP=0) \text{ OR } (\#JunctionBox_UP=0) \text{ OR } (\#InfraTI_UP=0) \text{ OR } (\#APP_UP=0)$
<i>Repair</i>	$((\#ACSource_UP>0) \text{ OR } (\#Generator_UP>0)) \text{ AND } (\#ATS_UP>0) \text{ AND } (\#VoltagePanel_UP>0) \text{ AND } (\#UPS_UP>0) \text{ AND } (\#Transformer_UP>0) \text{ AND } (\#SubPanel_UP>0) \text{ AND } (\#JunctionBox_UP>0) \text{ AND } (\#InfraTI_UP>0) \text{ AND } (\#APP_UP>0)$

Para calcular a disponibilidade do aplicativo usar a fórmula: $P\#ON > 0$.

2.4. Resultados

Esta seção apresenta os resultados dos modelos integrados. Foram realizadas análises de acordo com o uso da ferramenta Mercury. Considerou-se como parâmetros (Tabela 9): arquitetura de energia, mecanismos de checkpoint e tamanho do checkpoint; e como métricas: disponibilidade.

Tabela 9. Parâmetros e níveis

Parâmetros	níveis
Arquitetura de energia	A0, A1, A2 e A3
Mecanismo de Checkpoint	Cold, Warm e Hot
Tamanho do Checkpoint	1MB, 5MB, 10MB, 15MB, 20MB, 25MB e 100MB
Aplicação	VoD e Livraria Digital

Para avaliar o impacto do checkpoint nesses cenários, é encontrado na tabela 10 os valores de MTTR dos mecanismo para cada tamanho do checkpoint. Os valores foram retirados do [Gomes 2016].

Os resultados da métrica, disponibilidade, serão apresentados por tamanho do checkpoint. Sendo assim, na Figura 10 pode ser visto os valores da disponibilidade por

Tabela 10. MTTR mecanismo de checkpoint (em segundos) [Gomes 2016]

Mecanismo	1MB	5MB	10MB	15MB	20MB	25MB	100MB
Cold	0.3955	0.7415	0.862	1.05261	1.22309	1.3695	4.321
Warm	0.383	0.691	0.8305	0.94281	1.1079	1.2715	3.8621
Hot	0.2515	0.255	0.254	0.25699	0.26209	0.2515	0.2677

mecanismo de checkpoint. E para cada mecanismo, os valores de cada arquitetura de energia. Na figura 10(a) são apresentados os valores da aplicação VoD e na Figura 10(b) da aplicação de Livraria Digital.

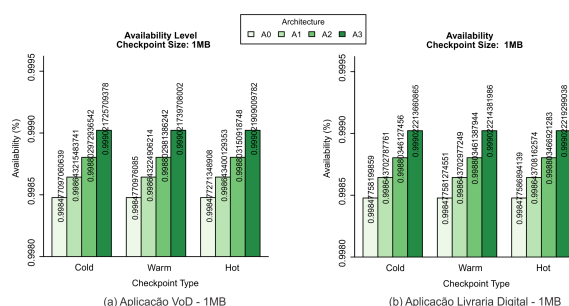


Figura 10. Disponibilidade - 1MB

Nas tabelas 11 e 12 são descritos os valores para checkpoint de tamanho 1MB, para VoD e Livraria Digital, respectivamente.

Tabela 11. Disponibilidade (em %) - Aplicação VoD - 1MB

Mecanismo/arquitetura	A0	A1	A2	A3
Cold	99.84770971	99.86432155	99.88029729	99.90217257
Warm	99.84770976	99.86432249	99.88029814	99.90217397
Hot	99.84772713	99.86434001	99.88031509	99.9021909

Tabela 12. Disponibilidade (em %) - Aplicação Livraria Digital - 1MB

Mecanismo/arquitetura	A0	A1	A2	A3
Cold	99.8477582	99.86437028	99.88034613	99.90222137
Warm	99.84775813	99.8643703	99.88034614	99.90222144
Hot	99.84775869	99.86437082	99.88034669	99.90222193

Os valores são muito próximos, independente do mecanismo de checkpoint de escolha. O que altera é a arquitetura de energia escolhida, como esperado a arquitetura com maior nível de redundância (A3) obtêm os melhores resultados.

Na Figura 11 e Tabelas 13 e 14 são apresentados os valores para 100MB.

Os valores para 5MB, 10MB, 15MB, 20MB e 25MB apresentaram resultados próximos aos citados acima.

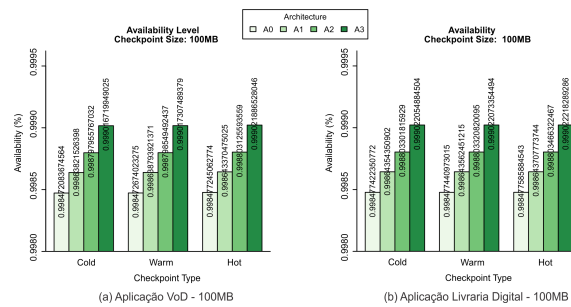


Figura 11. Disponibilidade - 100MB

Tabela 13. Disponibilidade (em %) - Aplicação VoD - 100MB

Mecanismo/arquitetura	A0	A1	A2	A3
Cold	99.84720837	99.86382153	99.87979558	99.90167199
Warm	99.8472674	99.86387939	99.87985495	99.90173075
Hot	99.84772451	99.86433705	99.88031256	99.90218865

Tabela 14. Disponibilidade (em %) - Aplicação Livraria Digital - 100MB

Mecanismo/arquitetura	A0	A1	A2	A3
Cold	99.84774224	99.86435435	99.88033018	99.90220549
Warm	99.8477441	99.86435625	99.88033208	99.90220734
Hot	99.84775859	99.86437078	99.88034663	99.90222183

Como pode-se notar, a arquitetura A3 obteve o melhor desempenho, e independente do mecanismo de checkpoint e do tamanho do checkpoint os valores de disponibilidade são poucos alterados.

Na aplicação de VoD, o menor valor de disponibilidade, 99.84720837%, foi apresentado quando se utilizou a arquitetura A0 com mecanismo de checkpoint Cold e com tamanho checkpoint de 100MB (ver Tabela 13). Já a disponibilidade mais alta, 99.9021909%, para essa aplicação foi encontrada na Arquitetura A3, utilizando checkpoint Hot com 1MB de tamanho (ver Tabela 11). A arquitetura A3, mesmo com nível elevado de redundância, obteve um aumento de, aproximadamente, 0.055% em relação a arquitetura A0.

Para a aplicação de Livraria Digital os valores foram muito próximos ao de VoD: variando de 99.84774224%, à cerca de 99.90222193% (Tabelas 14 e 12, respectivamente). O aumento da arquitetura A3 perante A0 foi de 0.0546%, aproximadamente.

3. Conclusão e próximos passos

Com uma maior dependência dos serviços de Internet, a disponibilidade do data center tornou-se uma preocupação mais séria. Para as empresas que dependem fortemente da Internet para suas operações, as interrupções do serviço podem ser muito caras, facilmente executadas em milhões de dólares por hora. Por isso, é de suma importância a compreensão de como um data center oferece disponibilidade para que, desta forma,

seja possível encontrar soluções mais sustentáveis e que não resulte em perda de desempenho.

O objetivo deste trabalho foi analisar o impacto da infraestrutura de energia na disponibilidade de aplicativos na nuvem, entender como se comporta um data center foi fundamental, comportamento este descrito na seção ???. Com base nisso, foram propostas quatro arquiteturas de energia, da mais simples, sem redundância, até a mais complexa, com maior número de componentes replicados. Vale ressaltar que as evoluções feitas de uma arquitetura para outra foram baseadas em análise de sensibilidade. Foi utilizado duas aplicações diferentes: VoD e Livraria Digital.

O modelo SPN foi utilizado para modelar as infraestruturas de energia e o mecanismo de checkpoint. Um modelo genérico para representar as falhas dos equipamentos de TI também foi realizado usando o RBD. As análises foram feitas na ferramenta Mercury.

Analizou-se o impacto das infraestruturas de energia na aplicação em vários cenários. Os cenários variam de quatro arquiteturas de energia, dois tipos de aplicação, três modos de checkpoint e sete tamanhos de checkpoint. Observou-se que o uso do checkpoint equivale a um ganho alto na disponibilidade, porém a diferença entre os modos de mecanismos de checkpoint é muito suave.

Referências

- Araujo, J., Maciel, P., Torquato, M., Callou, G., and Andrade, E. (2014). Availability evaluation of digital library cloud services. In *Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on*, pages 666–671. IEEE.
- Callou, G., Maciel, P., Tutsch, D., Ferreira, J., Araújo, J., and Souza, R. (2013). Estimating sustainability impact of high dependable data centers: a comparative study between brazilian and us energy mixes. *Computing*, 95(12):1137–1170.
- Cepin, M. (2011). *Assessment of Power System Reliability: Methods and Applications*. Springer London.
- Endo, P. T., de Almeida Palhares, A. V., Pereira, N. N., Goncalves, G. E., Sadok, D., Kelner, J., Melander, B., and Mangs, J.-E. (2011). Resource allocation for distributed cloud: concepts and research challenges. *IEEE network*, 25(4).
- Gomes, D. (2016). Performance evaluation of checkpoint services for multi-tier stateful applications.
- Melo, R., Bezerra, M. C., Dantas, J., Matos, R., Melo, I., and Maciel, P. (2015). Video on demand hosted in private cloud: Availability modeling and sensitivity analysis. In *Dependable Systems and Networks Workshops (DSN-W), 2015 IEEE International Conference on*, pages 12–18. IEEE.
- Reisig, W. (2013). *Understanding Petri Nets*. Springer, 1 edition.
- Singh, D., Singh, J., and Chhabra, A. (2012). High availability of clouds: Failover strategies for cloud computing using integrated checkpointing algorithms. In *Communication Systems and Network Technologies (CSNT), 2012 International Conference on*, pages 698–703. IEEE.

Mineração de Dados no Censo Escolar em Pernambuco

Wedson Lino Soares¹, Roberta A. de A. Fagundes^{1 2}

¹Universidade de Pernambuco, Campus Gov. Miguel Arraes de Alencar, Polo Comercial, \BR 104, Km 62 Caruaru (PE)- Brasil

² Departamento de Engenharia da Computação – Universidade de Pernambuco,
R. Benfica, 455 - Madalena, CEP:50720-001 - Recife - (PE) – Brasil

wedsonlino27@gmail.com, roberta.fagundes@upe.br

Abstract. *The school census is an important tool in improving the quality of education for the formulation, implementation and monitoring of educational policies and assessment of the performance of education systems. Thus, the article proposed aims to extract useful knowledge to identify the factors that influence school dropout on high school in public schools in Pernambuco. For the development of this article the R platform will be used during the study and will be the basis for creating an experimental environment for extracting useful information and evaluation of the proposed models in data mining phase using the CRISP- DM methodology. Currently, this work is concluded.*

Resumo. *O Censo Escolar é instrumento importante na melhoria da qualidade na educação para a formulação, implementação e monitoramento das políticas educacionais e avaliação do desempenho dos sistemas de ensino. Assim, o artigo proposto tem como objetivo extrair conhecimentos úteis para identificar os fatores que influenciam a evasão escolar no ensino médio das escolas públicas de Pernambuco. Para o desenvolvimento desse artigo a plataforma R será utilizada durante o estudo e servirá de base para criação de um ambiente experimental para extração de informações uteis e avaliação dos modelos propostos na fase de mineração de dados utilizando a metodologia CRISP-DM. Atualmente, esse trabalho encontra-se concluído.*

1.Introdução

O censo escolar[Inep 2015], realiza o levantamento de diversas informações estatístico-educacionais, relativas à educação básica em diferentes níveis (educação infantil, ensino fundamental e ensino médio). O levantamento é realizado, junto a todos os estabelecimentos de ensino, das redes pública e particular, através do preenchimento de

um questionário. Através do censo escola, o Inep pode atualizar diversas informações relativas às mais diversas áreas da educação.

Desta maneira, o censo escolar gera um conjunto de informações que são indispensáveis para a formulação, implementação e monitoramento de diversas políticas educacionais e avaliação do desempenho dos sistemas de ensino. Por isso o censo é tão importante, a partir dele é possível obter informações sobre as qualidades e deficiências do sistema educacional. Dentre as deficiências podemos citar: a evasão a escolar e a repetência, que combinados, são identificados, como sendo umas das principais falhas do sistema educacional brasileiro.

A evasão escolar[2009 Silva], é um problema que para ser solucionado, requer a intervenção de diversos, integrantes do meio acadêmico. É um problema que atinge países tanto ricos como pobres, e por isso vêm sendo foco de esforços, para tentar identificar, os indivíduos mais suscetíveis à evadir, através da análise de características pessoais e familiares, ou pela verificação de fatores relacionados às escolas nas quais estão matriculados.

Sendo assim, para analisar grandes bases de dados, como o censo escolar, a utilização de ferramentas de tecnologia da informação que auxiliam nesse processo tornou-se necessária. Para se tomar decisões de maneira, segura, inteligente e confiável, as organizações analisam de forma precisa, grandes volumes de dados, para minimização, de resultados não satisfatórios. Por isso tornou-se necessário, capturar os dados de forma precisa, descobrir o que é significativo e extrair informações relevantes. Sendo assim, matemática e estatística, tornam-se cada vez mais necessárias, para manipulação dos dados, e elaboração de modelos preditivos cada vez mais complexos.

Sendo assim, o objetivo desta pesquisa se concentra na aplicação de técnicas de mineração de dados à base de dados de escolas do censo escolar[Inep 2015] e à base das taxas de rendimento das Unidades federativas[Inep 2015] obtidas no site do Inep, usando como ferramenta a linguagem R [Software R], com foco na obtenção de características e análise de fatores que possam influenciar na evasão escolar do ensino médio das escolas de Pernambuco.

2. Introdução aos métodos de regressão

Nesta seção serão apresentados de maneira introdutória, alguns dos métodos que serão aplicados para a extração de informações do conjunto de dados do Censo escolar de 2015.

Verificar se duas ou mais variáveis estão relacionadas de alguma maneira é de grande interesse em diversas áreas do conhecimento. Modelos matemáticos podem ser empregados para expressar essa relação. Tal modelo é chamado de regressão, e ajuda a entender como a maneira que uma variável se comporta pode influenciar no comportamento de outra. A técnica abordada foi a regressão linear.

2.1. Regressão Linear

Na área de aprendizagem de máquina, a regressão linear é um modelo estatístico dos mais antigos. Os métodos supervisionados baseados neste tipo de regressão tentam modelar a relação entre entradas (variáveis explicativas) e saídas (variável resposta) na forma de equações paramétricas, onde os parâmetros são estimados a partir do método dos mínimos quadrados [Helene, 2006]. Este método minimiza a soma de quadrados das distâncias entre os pontos do diagrama e os respectivos pontos na curva da equação, tentando obter uma relação funcional entre X e Y para o modelo escolhido com o menor erro possível.

O modelo de regressão linear simples pode ser descrito da seguinte forma:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

A configuração acima representa o modelo de regressão linear simples. A variável X_i é a variável independente da equação enquanto Y_i é a variável dependente das variações de X , α é uma constante, que representa a interceptação da reta com o eixo vertical e β é outra constante, que representa o declive da reta.

Caso o modelo envolva uma relação entre duas variáveis, ele é chamado de simples. Quando envolver uma relação entre mais de duas variáveis ele é chamado de multivariado. Ou seja o comportamento de Y pode ser explicado por mais de uma variável independente $X_1, X_2, \dots, X_n$.

3. Evasão escolar

A evasão escolar[2], é um problema que para ser solucionado, requer a intervenção de diversos integrantes do meio acadêmico. É um problema que atinge países tanto ricos como pobres, e por isso vêm sendo foco de esforços, para tentar identificar, os indivíduos mais suscetíveis à evadir, através da análise de características pessoais e familiares, ou pela verificação de fatores relacionados às escolas nas quais estão matriculados.

Sendo assim, para analisar grandes bases de dados, como o censo escolar, a utilização de ferramentas de tecnologia da informação que auxiliam nesse processo tornou-se necessária. Para se tomar decisões de maneira, segura, inteligente e confiável, as organizações analisam de forma precisa, grandes volumes de dados, para minimização, de resultados não satisfatórios. Por isso tornou-se necessário, capturar os dados de forma precisa, descobrir o que é significativo e extrair informações relevantes. Sendo assim, matemática e estatística, tornam-se cada vez mais necessárias, para manipulação dos dados, e elaboração de modelos preditivos cada vez mais complexos.

3.1 Conjunto de dados: Censo Escolar 2015 [Inep] e Taxas de rendimento por UFs [Inep]

O conjunto de dados utilizados foram o Censo escolar de 2015, e as taxas de rendimento por Ufs também de 2015. Ambos fornecem informações sobre os diversos aspectos relacionados às condições da educação nas escolas brasileiras. Por intermédio do Censo Escolar, o Inep atualiza anualmente o Cadastro Nacional de Escolas e as informações referentes à matrícula, ao movimento e ao rendimento dos alunos, incluindo dados sobre sexo, turnos, turmas, séries e períodos, condições físicas dos prédios escolares e equipamentos existentes, além de informações sobre o pessoal técnico e administrativo e pessoal docente, por nível de atuação e grau de formação. Já o conjunto de dados das taxas de rendimento por UF fornece informações sobre abandono, aprovação e reprovação nas diversas redes de ensino.

Tabela 1. variáveis após o pré-processamento

TP_DEPENDENCIA	Dependência administrativa da escola (federal, municipal, estadual ou privada)
TP_LOCALIZACAO	Localização da escola (urbana ou rural)
TP_ATIVIDADE_COMPLEMENTAR	Atividade complementar (não oferece, não oferece exclusivamente ou oferece exclusivamente)
ABANDONO_1	Porcentagem de abandono no 1º ano do ensino médio nas escolas estaduais (retirado do conjunto de dados taxas de rendimento por UF)
ABANDONO_2	Porcentagem de abandono no 2º ano do ensino médio nas escolas estaduais (retirado do conjunto de dados taxas de rendimento por UF)
ABANDONO_3	Porcentagem de abandono no 3º ano do ensino médio nas escolas estaduais (retirado do conjunto de dados taxas de rendimento por UF)

A fim de estabelecer uma aplicação que retornasse estimativas de quais variáveis influenciam mais na evasão escolar, atributos do conjunto de dados citado acima foram utilizados para estabelecimento dos modelos de regressão. Para selecionar quais variáveis seriam utilizadas foi usado o critério de correlação, que mede o grau de associação entre duas variáveis.

4. Metodologia

A metodologia utilizada durante esta pesquisa, foi a CRISP-DM (CRoss Industry Standard

Process for Data Mining)[Reinartz 1998], que consiste em 6 etapas:

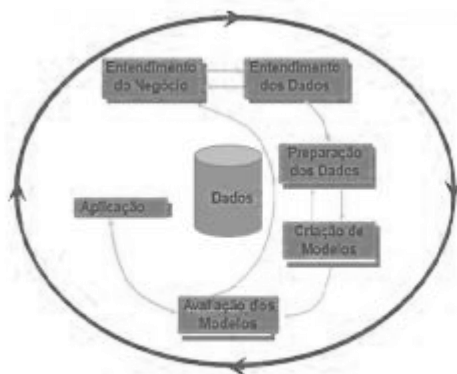


Figura 1 : CRISP – DM – Fonte: próprio autor

1. Conhecimento do negócio: é a etapa de conhecimento do problema, onde são definidos os objetivos e critérios para a interpretação dos resultados. O problema proposto consiste em encontrar possíveis fatores nas escolas que influenciem na evasão escolar. Nesta pesquisa, foram escolhidas para serem utilizadas a base de dados de escolas obtidas a partir do censo escolar de 2015, e a base de dados de taxas de rendimentos por UF, ambas obtidas a partir do site do Inep.
2. Compreensão dos dados: é a etapa que consiste na apreensão de conhecimento acerca dos dados que serão trabalhados mais à frente na mineração. Nesta pesquisa Foram usadas as bases de dados de escolas do censo escolar de 2015 e as bases de dados de taxas de rendimento por UF de 2015.
3. Preparação dos dados: primeiramente foi observado que existiam atributos cujo valor era do tipo binário ou cujas linhas na base de dados eram predominantemente vazias este tipo de dado não serviria para a aplicação da técnica selecionada, então foram eliminados.
4. Modelagem: é a etapa que consiste na aplicação das técnicas de mineração escolhidas para a obtenção dos resultados. Esta pesquisa teve como finalidade a aplicação de 2 diferentes técnicas de seleção de variáveis, seguidas da aplicação de técnicas de regressão linear por meio da função `lm()`, que recebe como primeiro parâmetro uma variável dependente, e uma ou mais variáveis independentes, a função então retorna como resultado o nível de relação entre essas variáveis. Para a seleção de variáveis foram utilizados dois métodos:
 - A seleção manual, observando a maior correlação entre as variáveis de abandono e as demais.
 - Seleção por meio da função `step`, que compreende um método para seleção automática de variáveis. Os parâmetros dessa função são uma função `lm`, esta por sua vez com seus próprios parâmetros, e a

indicação da direção que a função deverá seguir, que neste caso, foi utilizado o backward, no qual a função irá escolher as melhores variáveis para se ter em um modelo de regressão, por meio de eliminação.

Após a seleção das melhores variáveis para se ter em um modelo de regressão, obtidas por meio dos métodos citados acima, foi aplicada a função `lm()` do R, usando as variáveis indicadas como independentes, e as variáveis: `ABANDONO_1`, `ABANDONO_2` e `ABANDONO_3` como sendo as variáveis dependentes, para que fosse possível visualizar a influência das variáveis independentes sobre a evasão.

Logo após foi realizada uma comparação dos resultados obtidos para averiguação da integridade da informação. A aplicação das técnicas de mineração foi efetuada por meio do software R

5. Avaliação: nesta etapa os resultados obtidos são avaliados, Pois caso haja algum erro, é necessário voltar para a primeira etapa. Onde será realizada uma reavaliação sobre a base e buscar as devidas soluções para o problema encontrado. No caso desta pesquisa, os resultados foram obtidos de maneira esperada, sendo assim não foi necessário retornar à primeira etapa do CRISP-DM.
6. Implantação: é a etapa responsável pela entrega dos resultados finais obtidos. Os resultados podem ser conferidos a seguir.

5. avaliação experimental

Após a preparação das base de dados e do cruzamento das bases, foi aplicada a função `cor()` do R, que recebe como parâmetro duas variáveis, e aplica o método de correlação de Pearson[8] entre as duas para a observação do comportamento de uma variável em função da outra, quanto mais próximo a -1, este valor indica que uma variável decresce em função da outra. Enquanto que quanto mais próximo a 1, o valor revela que uma variável cresce em função da outra. Esta função foi aplicada para a obtenção de uma matriz de correlações, para uma melhor visualização das relações entre as variáveis de abandono e as demais variáveis.

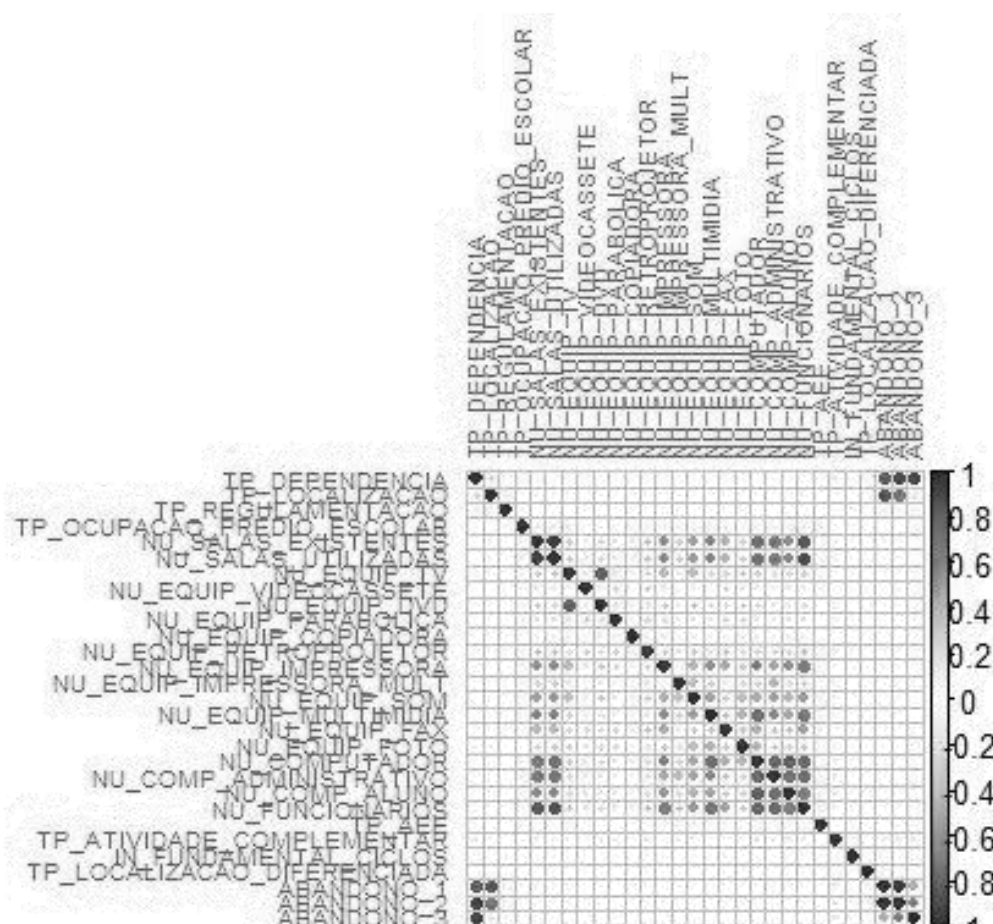


Figura 2 : nível de correlação entre as variáveis – Fonte: Próprio autor

Utilizando essa matriz, como ponto de partida, observa-se que as variáveis de ABANDONO_1, ABANDONO_2 e ABANDONO_3, tem uma correlação mais forte Com as seguintes variáveis: TP_DEPENDENCIA, TP_LOCALIZACAO e TP_ATIVIDADE_COMPLEMENTAR, que respectivamente significam a dependência administrativa (federal, estadual, municipal ou privada) a localização (rural ou urbana) das escolas e se ela oferece alguma atividade complementar de maneira exclusiva, não exclusiva ou se ela não oferece nenhuma atividade complementar. A partir dessas variáveis, foi estabelecida um modelo de regressão linear usando a função lm do R.

Do resultado obtido da função lm foi possível observar que ele indicou como as variáveis consideradas mais significantes para uma regressão linear, as variáveis, TP_DEPENDENCIA e TP_LOCALIZACAO. A função também mostrou que as variáveis exibiam um p-value bem baixo, com valor $< 2e-16$ e um valor de F-statistic > 1 , usando esses dois itens como indicadores, podemos rejeitar a hipótese nula de que as variáveis independentes não tem nenhuma relação com a variável dependente, e concluir que as variáveis indicadas pela função se relacionam com as variáveis de abandono.

Para uma maior precisão, foi aplicado a mesma função, usando-se apenas uma das variáveis indicadas anteriormente como tendo relação forte com a variável dependente de cada vez, para analisar se o valor do F-statistic seria maior do que 1 para assim, se ter certeza de que se poderia rejeitar a hipótese nula em todos os casos. Os resultados obtidos foram:

Tabela 2. resultados do F-statistic

	TP_LOCALIZACAO	TP_DEPENDENCIA
ABANDONO_1	3473	2781
ABANDONO_2	1512	6902
ABANDONO_3	66.44	7113

Sendo assim pode-se observar que em todos os casos, usando o F-statistic como indicador, é permitido rejeitar a hipótese nula de que as variáveis, não possuem nenhuma relação com as variáveis dependentes, no caso as variáveis de abandono.

Considerando que no método descrito acima, as variáveis a serem usadas na função lm foram selecionadas manualmente, também foi feito um teste no qual eram passadas todas as variáveis como sendo independentes, e as variáveis de abandono como variáveis dependentes, para avaliar se o resultado diferenciaria da seleção manual.

Foi observado que passando todas as variáveis como sendo independentes e a variável de abandono como variável dependente, o algoritmo escolheu como sendo as mais propícias para se estabelecer um modelo de regressão as variáveis TP_DEPENDENCIA e TP_LOCALIZACAO. Esse resultado indica que a seleção manual, e uma seleção mais automatizada acabou por apontar para duas das mesmas variáveis. novamente o valor de p-value das variáveis foi $< 2e-16$ e o valor de F-statistic > 1 , indicando a existência de relação entre tais variáveis e as variáveis de abandono.

Para fins de observação, também foi realizado um teste usando a função step do R, que basicamente é um meio para seleção automática de variáveis. Passando-se todas as variáveis como sendo as variáveis independentes, usando as variáveis de abandono como variáveis dependentes, e usando o método de seleção por eliminação. Observou-se que a função step terminou por apontar para aquelas mesmas variáveis observadas anteriormente. São elas as variáveis TP_DEPENDENCIA e TP_LOCALIZACAO.

6. Conclusão

Baseando-se no fato de que os três diferentes métodos aplicados, terminaram por apontar para as mesmas variáveis, como sendo variáveis que mantém relação forte com as variáveis de abandono, adicionado ao fato de que os indicadores de confiabilidade da regressão linear acabaram por reforçar esse resultado, pode-se utilizar isto como guia, e concluir que o fato da escola de um estudante do ensino médio ser federal, municipal, estadual ou privada, e se fica localizada em zona rural ou urbana, acabam por serem fatores decisivos para que este estudante venha a cometer a evasão escolar.

Os resultados obtidos, permitiram uma observação mais profunda sobre a influência de determinadas variáveis sobre a evasão escolar. Bem como o aprendizado sobre diferentes técnicas de mineração de dados, e sobre qual a melhor técnica utilizar quando estiver observando determinado tipo de problema.

7. Referências

Portal do Inep,

<http://portal.inep.gov.br/censo-escolar>

R software,

<https://cran.r-project.org>

Silva, L. C. F. da (2009). Evasão Escolar: Fatores Associados e Boas Práticas de Prevenção e Remediação.

Helene, O. (2006). Métodos dos Mínimos Quadrados, Editora Livraria da Física.

Reinartz, T.; Wirth, R.; Clinton, J.; Khabaza, T.; Hejlesen, J.; Chapaman, P.; Kerber, R. (1998). The current CRISP-DM process model for data mining. Berlin: Technical University.

Filho, F.; Brito, D.; Junior, S; Alexandre, J. (2010). Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r).Revista Política Hoje, v. 18, n. 1.

Estudo Experimental com as Ferramentas Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner*

Breno G. de Luna¹, Matheus F. X. Barboza¹, Paulemir G. Campos¹

¹ Universidade de Pernambuco (UPE) – Campus Caruaru
Caruaru – PE – Brasil

breno.luna_@hotmail.com, matheusreivax@gmail.com, pgc@poli.br

Abstract. *In this paper is presented an experimental study with Data Mining (DM) tools, such as, Oracle Data Mining, IBM SPSS and SAS Enterprise Miner. It is goal to discover at that problems may be used, including, to verify if can treat Big Data, these tools were applied to three databases. Additionally, it is also with this research aim to decision making by organizational managers, helping them to choose tools from DM more appropriate to their daily work.*

Resumo. *Neste artigo é apresentado um estudo experimental com as ferramentas de Mineração de Dados (MD) Oracle Data Mining, IBM SPSS e SAS Enterprise Miner. Almejando descobrir em que problemas poderão ser utilizadas, inclusive, verificar se tratam Big Data, essas ferramentas foram aplicadas a três bases de dados. Adicionalmente, pretende-se também com essa pesquisa ajudar a tomada de decisão por gestores organizacionais, auxiliando-os a escolher ferramentas de MD mais apropriadas ao seu cotidiano de trabalho.*

1. Introdução

Na década de 1970, pesquisadores e estudiosos da área de Inteligência Artificial (IA) verificaram que para desenvolver sistemas computacionais para resolver de forma adequada problemas do mundo real, deveria-se adicionar nesses sistemas grandes quantidades de conhecimento específico sobre o problema [Russel e Norvig 2013].

No entanto, com o uso de ferramentas automáticas para coleta de dados somado com a maturidade das tecnologias de bancos de dados e a existência de enormes volumes de dados provenientes de bases de dados e outros meios de armazenamento geraram uma verdadeira explosão de dados. Contudo, essa abundância de dados apresentava escassez de conhecimento, motivando desta forma a criação de uma área de pesquisa denominada Mineração de Dados (MD), vinculada à IA [Han et al. 2012].

Em torno dos últimos dez anos, esse grande interesse em desenvolvimento e proliferação de aplicações que usam *Cloud Computing* (Computação em Nuvem) desencadeou o aparecimento de *Big Data* [Silva e Campos 2015].

Ultimamente, [Sousa Neto e Lins Filho 2014] evidenciaram que:

O crescimento do volume de dados gerados a partir dos sistemas computacionais tem causado um grande impacto na área de TI. Com o advento do *Big Data* verifica-se que os recursos computacionais disponíveis não estão aptos a lidar com todos esses dados. [Sousa Neto e Lins Filho 2014, p. 1].

No mais, é relevante observar que em organizações do setor público ou privado, é clara a necessidade da aplicação de recursos tecnológicos satisfatoriamente, com o intuito de ter-se uma gestão organizacional eficiente e competitiva. Assim, e num mundo globalizado, é crescente a quantidade de organizações procurando utilizar Tecnologia da Informação (TI), Mineração de Dados (*Data Mining*), *Business Intelligence* (BI) e inovação tecnológica nos processos organizacionais e na tomada de decisão, nos mais diversificados ramos de atuação profissional, tais como, educação, comunicação, telecomunicações, saúde, mercado financeiro, administradoras de cartões de crédito, indústria, comércio, agricultura, pecuária, mineração e pesquisas espaciais [Laudon e Laudon 2010], [Witten et al. 2011], [Han et al. 2012], [Salgado et al. 2015], [Sousa Neto et al. 2015], [Silva e Campos 2015], [Mendonça 2015].

Portanto, neste artigo é apresentado um estudo experimental com algumas ferramentas de Mineração de Dados que possam auxiliar na tomada de decisão nos mais diversos domínios do conhecimento humano. Inclusive, destacando-se propriedades dessas ferramentas, tais como em que problemas podem ser aplicadas, sistema operacional exigido, tipo de licença de *software*, algoritmos de Mineração de Dados disponíveis, métodos de Visualização de Dados e até se podem tratar *Big Data*.

Finalmente, este artigo, além desta seção introdutória, contém: A segunda seção, descrevendo ferramentas de Mineração de Dados; A terceira seção aborda experimentos com ferramentas de Mineração de Dados; Já a quarta seção exhibe os resultados obtidos; E, na quinta e última seção, são mostradas as conclusões desta pesquisa.

2. Ferramentas de Mineração de Dados

A partir da necessidade de extrair conhecimento de grandes volumes de dados eficientemente, diversos institutos de pesquisa e empresas concentraram esforços no desenvolvimento de ferramentas de Mineração de Dados [Witten et al. 2011], [Han et al. 2012], [Sousa Neto et al. 2015].

Nos dias de hoje, existem várias ferramentas para ajudar a descoberta de conhecimento de grandes bases de dados em muitas áreas do conhecimento humano, almejando, assim, auxiliar a tomada de decisão pelos gestores de organizações científicas e/ou empresariais. Destacam-se, dentre ferramentas de Mineração de Dados, as seguintes: Oracle *Data Mining*¹, IBM SPSS², SAS *Enterprise Miner*³, TANAGRA [Rakotomalala

¹ Disponível em: <http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html>, acessada em junho de 2017.

2005], WEKA [Hall et al. 2009], Orange [Demsar et al. 2013], KNIME⁴, Miner3D⁵ e RapidMiner⁶.

As ferramentas WEKA, Orange e KNIME, todas estas com licença de *software* GNU (*General Public Licence*), foram investigadas por Melo et al. (2016). Por sua vez, dando continuidade a esse estudo, porém, com enfoque estritamente empresarial e considerando apenas ferramentas de Mineração de Dados com licença de *software* proprietário, serão analisadas de forma aprofundada as ferramentas Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner*.

2.1.Oracle *Data Mining*

O Oracle *Data Mining* (ODM) é um *software* muito poderoso, que provê mineração de dados, usando o Oracle *Database*. É possível usar o ODM para criar e implementar aplicações de mineração de dados preditiva e descritiva, para adicionar elementos de inteligência em aplicações já existentes e criar consultas preditivas para exploração de dados.

O ODM oferece, de forma compreensiva, vários algoritmos de banco de dados, que performam uma variedade de tarefas de mineração, como classificação, regressão, atribuição de importâncias, detecção de anomalia, criação de *clusters*, etc. Esses algoritmos podem trabalhar com uma base de dados padrão, arquivos transacionais, textos e outras formas de dados não estruturados. Esse *software* é mais adequado para a mineração de grandes bases de dados.

2.2. IBM SPSS

O IBM SPSS é um *software* que teve a sua primeira versão no ano de 1968 e foi criado por três pesquisadores americanos, Norman H. Nie, C. Hadlai Hull e Dale H. Bent na Universidade de Stanford⁷. Apenas em 1984 sua primeira versão para computadores pessoais foi lançada. O sistema foi baseado na ideia de usar estatísticas para transformar dados brutos em informações essenciais para a tomada de decisões.

O pacote de *software* IBM SPSS *Statistics* é uma das principais ferramentas para análise preditiva do mercado atual, possuindo uma grande variedade de *add-ons* e recursos para solucionar uma quantidade ampla de problemas relacionados a negócios e pesquisas que precisam de uma análise mais detalhada e precisa. O *software* utiliza e combina análise avançada e otimização de decisões, e continua a ter um foco para a organização, pois ela procura aumentar a interação do mercado dos benefícios empresariais que a análise preditiva pode oferecer.

² Disponível em: <http://www.ibm.com/analytics/us/en/technology/spss/>, acessada em junho de 2017.

³ Disponível em: http://www.sas.com/en_us/software/analytics/enterprise-miner.html, acessada em setembro de 2017.

⁴ Disponível em: <https://www.knime.org/>, acessada em julho de 2017.

⁵ Disponível em: <http://www.miner3d.com/>, acessada em agosto de 2017.

⁶ Disponível em: <http://docs.rapidminer.com/>, acessada em agosto de 2017.

⁷ Disponível em: <http://www.spss.com.hk/corpinfo/history.htm>, acessada em outubro de 2017.

2.3. SAS *Enterprise Miner*

O SAS *Enterprise Miner* é um *software* comercial de enorme potencial, feito exclusivamente pra se trabalhar com grandes bases de dados. Trabalha com uma interface gráfica bem simples, e resultados visuais, como gráficos, tabelas, etc. Suporta o uso de análises preditivas e descritivas, possuindo mais técnicas de modelagem do qualquer outro *software* no mercado, tornando-o um pouco mais complexo de usá-lo de forma 100% satisfatória.

Por se tratar de uma empresa com grande nome no mercado, a SAS *Enterprise* conta com mais de 80.000 clientes, nos quais ela os ajuda para realizarem tomadas de decisões, desde 1976. Por causa disso, ele vem se tornando cada vez mais importante no mercado de mineração de dados e análises, e possui uma grande equipe que dá o devido suporte para seus usuários.

3. Experimentos com Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner*

Os experimentos foram realizados com as ferramentas Oracle *Data Mining* 4.1, IBM SPSS 23 e SAS *Enterprise Miner* 14.2 com o Sistema Operacional Windows 10 de 64 bits, usando as bases de dados de Câncer de Mama⁸, de Íris⁹ e de Vinho¹⁰, obtidas do repositório de Aprendizado de Máquina da Universidade da Califórnia [Lichman 2013].

Nesses estudos experimentais, verificou-se funcionalidades de pré-processamento dos dados, algoritmos de Mineração de Dados - principalmente aplicados a problemas de classificação, de regressão e de agrupamento (*clustering*), bem como métodos de visualização de dados, de avaliação do conhecimento extraído e se poderiam lidar com *Big Data*.

É válido ressaltar que em todos os algoritmos de Mineração de Dados, exceto nos algoritmos de agrupamento onde necessitou-se configurar a quantidade de *clusters* convenientemente, os parâmetros usados foram os valores padrões (*default*) predefinidos em cada uma das ferramentas de Mineração de Dados investigadas.

4. Resultados

Depois da execução dos experimentos com as ferramentas Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner* foram encontrados os resultados ilustrados nas Figuras 1, 2 e 3, e na Tabela 1 a seguir.

⁸ Disponível em: <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>, acessada em julho de 2017.

⁹ Disponível em: <http://archive.ics.uci.edu/ml/machine-learning-databases/iris/>, acessada em julho de 2017.

¹⁰ Disponível em: <http://archive.ics.uci.edu/ml/machine-learning-databases/wine/>, acessada em julho de 2017.

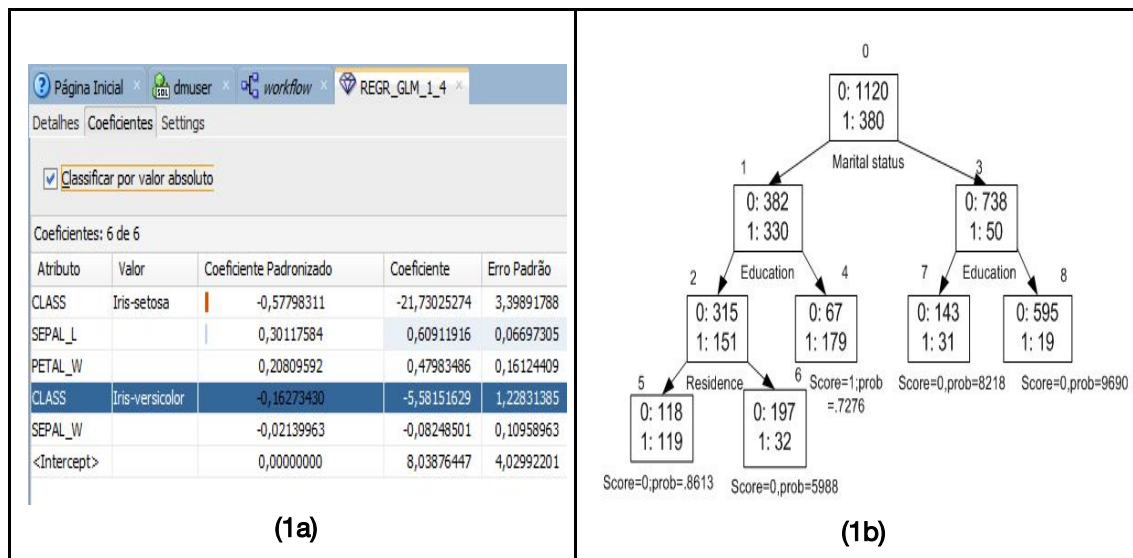


Figura 1. Resultados gráficos obtidos com a Ferramenta Oracle Data Mining.
(1a) Usando Regressão Linear. (1b) Usando Árvore de Decisão

Na Figura 1 é aplicado um modelo matemático de Regressão Linear disponível na ferramenta Oracle *Data Mining* para se obter dados preditivos (Figura 1a), e um algoritmo de Árvore de Decisão (Figura 1b) num problema de classificação.

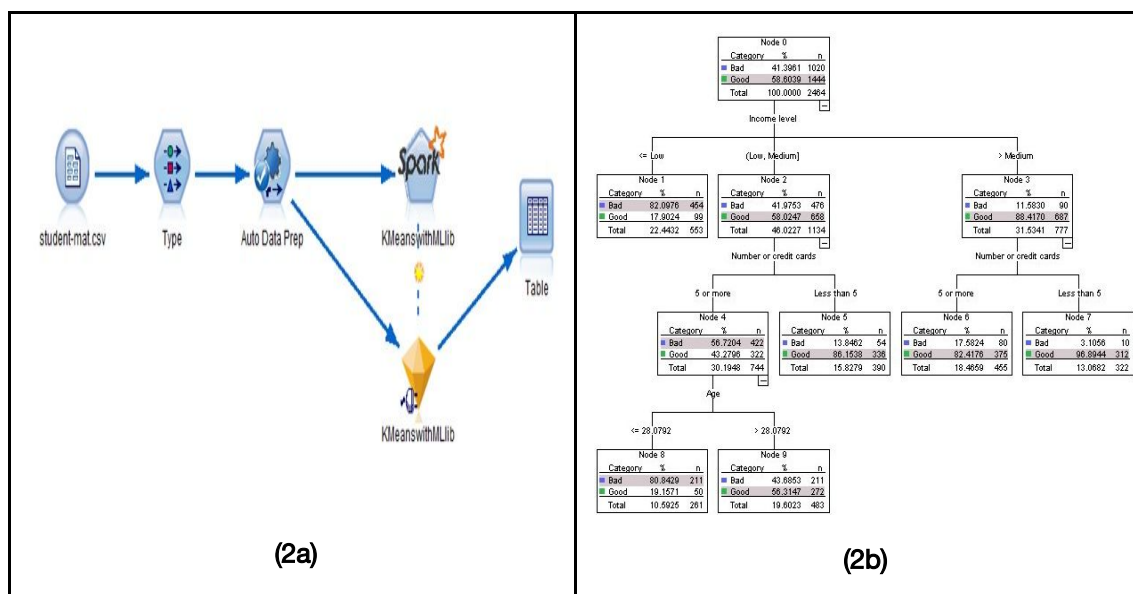


Figura 2. Resultados gráficos construídos com a Ferramenta IBM SPSS.
(2a) Usando Clustering. (2b) Usando Árvore de Decisão.

Já na Figura 2 são ilustrados usando a ferramenta de mineração de dados IBM SPSS a aplicação de um algoritmo de agrupamento (*clustering*, Figura 2a) e de uma Árvore de Decisão (Figura 2b).

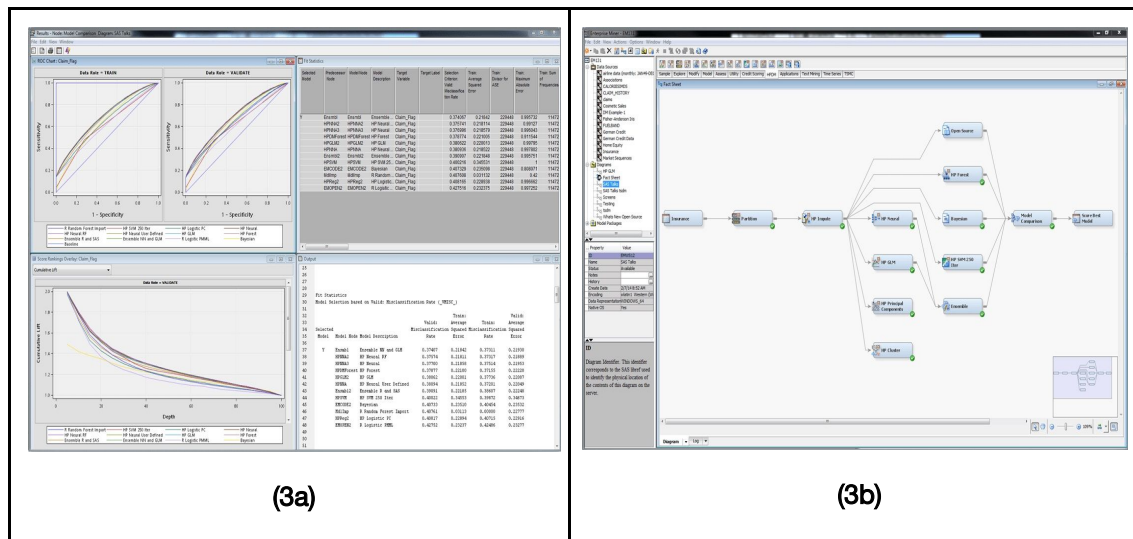


Figura 3. Resultados gráficos construídos com a Ferramenta SAS Enterprise Miner.
(3a) Comparação de Modelos Distintos. (3b) Diagrama de Fluxo de Processo.

Por sua vez, a Figura 3¹¹ exemplifica recursos oferecidos pela ferramenta SAS Enterprise Miner, como a comparação de vários modelos distintos simultaneamente (Figura 3a), e a elaboração de um diagrama de fluxo de processo (Figura 3b).

Tabela 1. Propriedades das Ferramentas de Mineração de Dados Oracle Data Mining, IBM SPSS e SAS Enterprise Miner

Propriedades	Ferramentas de Mineração de Dados		
	Oracle Data Mining	IBM SPSS	SAS Enterprise Miner
Sistema Operacional	Windows, Unix e Linux.	Windows, Linux e Mac OS X.	Windows, Unix e Linux.
Licença	Proprietária, mas com versão de teste.	Proprietária, mas com versão de teste.	Proprietária
Suporte a Big Data	Sim.	Sim.	Sim.
Tipos de Problemas	Classificação, Regressão, Agrupamentos, Associação.	Classificação, Regressão, Agrupamentos, Previsão.	Classificação, Regressão.

¹¹ Disponível em: https://www.sas.com/en_us/software/enterprise-miner.html, acessada em outubro de 2017.

**Tabela 1 (Continuação). Propriedades das Ferramentas de Mineração de Dados
Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner***

Propriedades	Ferramentas de Mineração de Dados		
	Oracle <i>Data Mining</i>	IBM SPSS	SAS <i>Enterprise Miner</i>
Pré-Processamento dos Dados	Preparação de Dados Nominais e Numéricos, Importância de Atributos, Detecção de Anomalias, Extração e Seleção de Características.	<i>Missing values</i> , seleção de atributos, redução e transformação de dados.	<i>Missing values</i> , seleção de atributos, transformação de dados.
Árvores de Decisão	Sim.	Sim.	Sim.
Regras de Decisão	Não.	Não.	Não.
Redes Neurais	Sim.	Sim.	Sim.
Regressão	Modelos Lineares Generalizados de Regressão Múltipla e SVM (<i>Support Vector Machine</i>)	Regressão Linear e Regressão Logística.	Regressão Linear e Regressão Logística.
Meta-Aprendizado	Não.	Não.	<i>Boosting</i> .
Agrupamentos	K-Means, Maximização da Expectativa, Particionamento Ortogonal.	Sim.	Não.
Regras de Associação	Apriori.	Não.	Não.

**Tabela 1 (Continuação). Propriedades das Ferramentas de Mineração de Dados
Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner***

Propriedades	Ferramentas de Mineração de Dados		
	Oracle <i>Data Mining</i>	IBM SPSS	SAS <i>Enterprise Miner</i>
Estatísticos	<i>Naive Bayes.</i>	Estatística Descritiva e Bayesiana, Análise Fatorial, Teste de Hipótese.	Não.
Visualização dos Dados	Diversos Gráficos e Tabelas.	Diversos Gráficos, Tabelas, Relatórios e Diagramas.	Diagramas, Tabelas e Gráficos.
Avaliação dos Dados	<i>Holdout.</i>	<i>Bootstrap.</i>	<i>Holdout.</i>

E, a partir dos resultados exibidos na Tabela 1, verifica-se que:

- Essas ferramentas possuem boa interface gráfica, executam no Windows e Linux, podem ser aplicadas a problemas de Classificação e Regressão, contém métodos para pré-processamento dos dados, possuem um número limitado de algoritmos de Mineração de Dados, e diversas formas para Visualização dos Dados de saída;
- Todas as ferramentas analisadas podem trabalhar com *Big Data* por se tratarem de *softwares* comerciais usados por grandes empresas;
- Nenhuma dessas ferramentas suporta algoritmos de regras de decisão;
- E, só SAS *Enterprise Miner* não possui versão gratuita para testes.

5. Conclusões

Pode-se concluir, de acordo com os experimentos e resultados obtidos com as ferramentas Oracle *Data Mining*, IBM SPSS e SAS *Enterprise Miner*, que os objetivos dessa pesquisa foram atingidos, tornando a mineração de dados algo mais intuitivo e prático, com resultados claros, facilitando o ato de tomar decisões, pois agora é possível enxergar mais padrões e formas nos dados.

Futuramente, a pesquisa será ampliada, com um aprofundamento no uso dessas e de outras ferramentas, como CMSR *Data Miner*¹² e *Knowledge Studio*¹³, e de seus algoritmos, para se tirar maior proveito na hora de utilizá-los, podendo assim, identificar diferenças mais minuciosas na hora de se escolher uma ferramenta para se obter um maior desempenho naquilo que se deseja realizar para uma empresa.

Referências

- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M. e Zupan, B. (2013) “Orange: Data Mining Toolbox in Python”, *Journal of Machine Learning Research*, Volume 14, p. 2349-2353.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. e Witten, I. H. (2009) “The WEKA Data Mining Software: An Update”, *SIGKDD Explorations*, Volume 11, Issue 1.
- Han, J., Kamber, M. e Pei, J. (2012) *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 3rd ed., San Francisco.
- Laudon, K. C. e Laudon, J. P. (2010) *Sistemas de Informação Gerenciais*, Pearson Prentice Hall, 9a. ed., São Paulo.
- Lichman, M. (2013) *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>], School of Information and Computer Science, University of California, Irvine.
- Melo, A. B., Silva, G. A., Moraes, P. L. e Campos, P. G. (2016) Estudo Experimental com as Ferramentas de Mineração de Dados WEKA, Orange e KNIME. In *Anais da Jornada Científica e de Extensão - JCE 2016 da UPE Caruaru*, p. 55-62, Caruaru-PE.
- Mendonça, C. M. C. (2015) *A Computação em Nuvem e as suas Relações com os Arranjos de Governança de TI e o Alinhamento Estratégico*, Tese de Doutorado em Administração, Universidade Federal do Rio Grande do Norte, Natal-RN.
- Rakotomalala, R. (2005) “TANAGRA: Un Logiciel Gratuit pour L’Enseignement et la Recherche”, *Actes de EGC’2005, RNTI-E-3*, Volume 2, p. 697-702.
- Russel, S. e Norvig, P. (2013) *Inteligência Artificial*, Tradução de: “Artificial Intelligence: A Modern Approach”, Editora Campus, 3a. ed.
- Salgado, C. C. R., Araújo, A. G. e Sousa Neto, M. V. (2015) Tecnologia da Informação e Vantagem Competitiva: Exploração dos Artigos sob a Ótica da Bibliometria. In *XVIII SEMEAD’2015, Seminários em Administração*, 16p., Natal-RN.
- Silva, I. M. e Campos, F. C. (2015) Big Data e Cloud Computing Perspectivas de Uso em Engenharia de Produção: Uma Análise Bibliométrica de 2004-2014. In *XXXV ENEGEP’2015, Encontro Nacional de Engenharia de Produção*, 19p., Fortaleza-CE.

¹² Disponível em: <http://www.roselladb.com/starprobe.htm>, acessada em outubro de 2017.

¹³ Disponível em: <http://www.angoss.com/predictive-analytics-software/software/knowledgestudio/>, acessada em outubro de 2017.

- Sousa Neto, M. V. de e Lins Filho, M. L. (2014) Desafios na Era do Big Data: Uma Revisão. In *XXI SIMPEP'2014, Simpósio de Engenharia de Produção*, 14p., Bauru-SP.
- Sousa Neto, M. V. de, Medeiros Junior, J. V., Anez, M. E. M. e Moraes, E. A. (2015) Building Information Technology Capabilities: A Case Study of the Development of an Integrated Management System. In *International Conference of the System Dynamics Society'2015, Proceedings of the 33rd International Conference of the System Dynamics Society*, 12p., Cambridge-USA.
- Teixeira, M. F. (2013) Agrupamento e Visualização de Dados: Estudo e Implementações para a Ferramenta YADMT, Trabalho de Conclusão de Curso em Ciência da Computação, Universidade Estadual do Oeste do Paraná, Cascavel-PR.
- Witten, I. H., Frank, E. e Hall, M. A. (2011) Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 3rd ed.

Fatores que influenciam a nota das escolas no ENEM

Geovanne Oliveira Alves¹, Roberta Andrade de Araújo Fagundes¹

¹Universidade de Pernambuco, Campus Gov. Miguel Arraes de Alencar, Polo Comercial, BR 104, Km 62 Caruaru (PE)- Brasil

{geovanne.oliveira, roberta.fagundes}@upe.br

Resumo. *O Exame Nacional do Ensino Médio (ENEM) que, atualmente, serve como parâmetro de acesso às universidades públicas e privadas do Brasil, fornece várias informações que podem ser usadas para melhor avaliar o desempenho do ensino. Assim, esse artigo apresenta os fatores econômicos e sociais que influenciam o desempenho das escolas do estado de Pernambuco no ENEM. Foram identificadas algumas variáveis influentes utilizando técnicas de Data Mining no software R. Pretende-se analisar esses fatores utilizando o método holdout no âmbito da simulação de Monte Carlo.*

1. Introdução

O Enem, Exame Nacional do Ensino Médio, além de avaliar o conhecimento obtido até o Ensino Médio e servir como processo seletivo de grande parte das universidades públicas e privadas do país, é possível extrair informações importantes das escolas e dos participantes. Fatores que afetem o desempenho acadêmico dos estudantes, como as condições socioeconômicas, devem ser analisados para a avaliação das escolas. Entretanto, em uma análise que considera apenas a nota das provas objetivas (excluindo redação), os dados apontam uma discrepância que separa as escolas públicas e as escolas privadas. Em 2015, das 100 escolas com maior nota média no Enem, 97 são privadas. No universo de 1 mil escolas, somente 49 são da rede pública. No ano 2014, eram 93, e em 2013, 78. Esse cenário considera o cálculo da média aritmética das quatro provas objetivas (linguagens, matemática, ciências humanas e ciências da natureza) [Globo 2017]. Por fatores como esse, é consenso entre especialistas em educação que a nota do Enem por escolas não pode ser tomada como único critério para escolha de uma escola [Douglas 2003].

Os dados mostram que a diferença de financiamento direto na escola também reflete o desempenho médio dos estudantes. A comparação entre as mil escolas com as maiores médias na nota do Enem e as mil escolas com as médias mais baixas revela que, no primeiro grupo, 951 escolas são privadas, e as demais são escolas públicas de regime e orçamento diferenciados. Essa é a concentração mais alta de escolas privadas entre as melhores do país em pelo menos três anos. No Enem 2014, entre as mil melhores escolas, 93 eram públicas. No Enem 2013, 78 das mil melhores escolas eram da rede pública [Globo 2017].

Para o Inep, novamente os resultados das escolas públicas apontam a necessidade de reforma do ensino médio. “Temos contingente enorme de alunos que estão aprendendo muito pouco, currículo organizado de maneira tradicionalista e conservadora. Os resultados apontam para a necessidade da reforma do ensino médio”.

Pesquisas feitas em vários países mostram que esses fatores [Douglas 2003] são um problema universal que envolve, para sua solução, distintas intervenções de alunos, pais, professores, autoridades educacionais, formuladores de políticas públicas e lideranças políticas, empresariais e comunitárias [Portal do Inep 2016]. Não obstante as particularidades institucionais de cada nação.

Assim, para analisar grandes bases de dados, como o ENEM, requer a utilização de ferramentas da tecnologia da informação que auxiliam esse processo. Com o intuito de tomar decisões de forma inteligente, segura e confiável as organizações analisam de forma precisa, grandes volumes de dados visando minimização de riscos e resultados não satisfatórios. Nesse sentido utilizar ferramentas que possam extrair informações de uma base de dados em forma de conhecimento voltado para a tomada de decisão. Por isso, tornou-se necessário capturar dados de forma precisa, classificar, descobrir o que é significativo, enxergar padrões e extrair informações relevantes. Matemática e estatística tornam-se cada vez mais necessárias para manipular os dados, metodologia científica para ter embasamento na elaboração de hipóteses e em modelos preditivos de cenários cada vez mais complexos e até lógica de programação passa a ser valiosa para trabalhar de forma avançada com o grande volume de dados na forma digital (ex.: em ferramentas de estatística como R, SPSS) [PETER 2002].

A partir da análise de dados, é possível identificar quais fatores que tem um impacto maior sobre as notas das escolas, ferramentas de *Data Mining* foram utilizadas para análise dos dados nesse artigo. *Data Mining* consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis e, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados [Devmedia 2017]. Um algoritmo relacionado à *Data Mining* é a Regressão Linear permite uma previsão de certos dados de acordo com a série histórica, seguindo um modelo linear.

2. Conceitos Básicos

A seguir serão abordados conceitos básicos do tema abordado neste trabalho para um maior entendimento acerca do mesmo.

2.1. Regressão Linear

Regressão Linear clássica (RL) é um dos mais antigos modelos estatísticos preditivo na área de aprendizagem de máquina. Os métodos supervisionados baseados neste tipo de regressão tentam modelar a relação entre entradas (variáveis explicativas) e saídas (variável resposta) na forma de equações paramétricas, onde os parâmetros são estimados a partir do método dos mínimos quadrados [Helene, 2006]. Este método minimiza a soma de quadrados das distâncias entre os pontos do diagrama e os respectivos pontos na curva da equação, tentando obter uma relação funcional entre X e Y para o modelo escolhido com o menor erro possível.

O modelo de regressão linear simples pode ser descrito da seguinte forma:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Esta configuração representa o modelo de regressão linear simples. A variável X_i é a variável independente da equação enquanto Y_i é a variável dependente das variações de X , α é uma constante, que representa a interceptação da reta com o eixo vertical e β é outra constante, que representa o declive da reta.

O modelo de regressão é chamado de simples quando envolve uma relação causal entre duas variáveis. O modelo de regressão é multivariado quando envolve uma relação causal com mais de duas variáveis. Isto é, quando o comportamento de Y é explicado por mais de uma variável independente $X_1, X_2, \dots, X_n$.

2.2. Correlação

A correlação do termo é usada com funções estatísticas, para referir-se ao movimento de duas ou mais variáveis em consequência fornecida pela função em estudo. A correlação basicamente ocorre quando dois elementos tem harmonia em variação, esta harmonia é dependente, ou seja, estabilização de um depende da posição do outro [Enciclopédia Culturama 2017]. Esse método é utilizado quando precisa verificar se existe relação entre duas ou mais variáveis, quanto a força e a direção o que pode ser demonstrado em um gráfico de Dispersão, porém, existe um número que mede a relação conhecido como coeficiente de correlação. O coeficiente de correlação pode variar de $-1,00$ a $+1,00$, com um coeficiente de $+1$, indicando uma correlação linear positiva perfeita [Lorí Viali 2017].

Assim dadas duas amostras, uma da variável X e outra da variável Y , sendo N o tamanho total, o coeficiente de correlação amostral poderá ser calculado através da seguinte expressão:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Considerando as amostras X e Y , onde X assumindo valores individuais para $X_1, X_2, X_3, \dots, X_n$ e Y assumindo valores individuais $Y_1, Y_2, Y_3, \dots, Y_n$. Assim, após aplicar a fórmula é obtido o coeficiente de correlação que fornece o grau de relacionamento linear entre duas variáveis.

2.3. Stepwise (Backward)

No conjunto de variáveis independentes pode haver variáveis com o coeficiente de correlação tão baixo que não influencia significativamente as variáveis dependentes (saída). O método *stepwise* é usado para selecionar quais variáveis mais influenciam o conjunto de saída podendo, assim, diminuir o número de variáveis a compor a equação de regressão. Uma das modificações de seleção do *stepwise* é o método *backward*.

O método *Backward* inicia o processo com todas as variáveis auxiliares do modelo proposto e depois, por etapas, cada uma pode ou não ser eliminada [Portal Action 2017]. A decisão de retirada da variável é tomada baseando-se no método AIC, que são calculados para cada etapa, referentes àquele modelo. O Critério de Informação de Akaike (AIC) é definido como:

$$AIC_p = -2 \log(L_p) + 2[(p + 1) + 1]$$

Em que L_p é a função de máxima verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo.

3. Metodologia

A metodologia mais aceita em mineração de dados é conhecida por CRISP-DM [Reinartz et al. 1998], sigla do inglês *Cross-Industry Standard Process for Data Mining*. O uso da metodologia CRISP nos permite atacar o problema de extração do conhecimento de forma progressiva e organizada, partindo de uma análise de alto nível, que busca a compreensão do negócio, e indo em direção à definição e implantação de modelos que permitam efetivamente atingir os objetivos da mineração.

Essa metodologia é constituída de seis etapas: Entendimento do negócio, Entendimento dos dados, Preparação dos dados, Modelagem, Avaliação e Aplicação (figura 1).

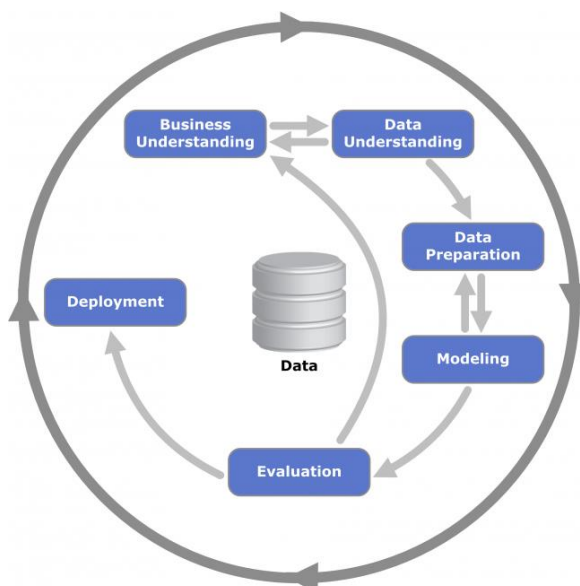


Figura 1. Metodologia CRISP-DM

Cada etapa é subdividida em tarefas definindo assim um processo de trabalho iterativo, ou seja, realizado em ciclos. Dessa forma, um artigo de mineração vai repetir o ciclo de etapas quantas vezes forem necessárias até que os objetivos da mineração tenham sido atingidos. Abaixo segue a discriminação de cada etapa dessa metodologia.

- Entendimento do negócio: Esse artigo busca identificar os fatores determinantes no desempenho das escolas do estado de Pernambuco através dos dados do ENEM, utilizando tarefas de *Data Mining*.
- Entendimento dos dados: Para esse estudo foi utilizado o questionário socioeconômico, juntamente com as notas de cada participante do ENEM 2015 para poder obter uma relação entre os dados. Na seção dos resultados foram listados as fontes, os procedimentos de leitura e os problemas detectados na descrição dos dados, assim como toda a informação relevante.
- Preparação dos dados: Esse processo foi constituído por várias etapas de filtragem e seleção dos dados na base em estudo. Assim foi possível juntar todos

os dados uteis em uma tabela, criando uma visão única de toda informação necessária para a análise.

- Modelagem: Foram aplicadas algumas técnicas de *Data Mining* sobre o conjunto de dados. Os métodos de regressão linear e correlação, assim como os procedimentos de escolha das variáveis utilizando o método *stepwise* foram essenciais para a escolha dos fatores mais influentes.
- Avaliação de desempenho: O desempenho do modelo será medido pelo erro médio relativo estimado pelo método Monte Carlo [Chernick 1999].
- Distribuição dos resultados: Nesse artigo estão discriminados os fatores sociais e econômicos que mais influenciam as notas das escolas no ENEM e os resultados encontrados até o momento.

4. Resultados

Essa seção apresenta os conjuntos de dados utilizados nesse estudo, especificando os procedimentos realizados na preparação dos dados, e os fatores sociais e econômicos mais representativos da população, descrevendo os processos estatísticos utilizados.

4.1. Conjunto de Dados

Para o estudo foi utilizada a base de micro dados do Enem 2015 [Microdados 2017] onde contém os dados de cada participante do Exame Nacional do Ensino Médio em 2015. Os micro dados do Inep se constituem no menor nível de desagregação de dados recolhidos por pesquisas, avaliações e exames realizados [Microdados 2017]. A base do ENEM contém os dados de um pouco mais de 7 milhões de candidatos de todo o Brasil, com o intuito de sintetizar a base para a análise mais objetiva foram necessários alguns procedimentos de filtragem dos dados.

Como o objetivo desse projeto visa as escolas do estado de Pernambuco, foi necessário uma filtragem para a seleção dos participantes do estado de Pernambuco, pois a base abrange os participantes de todo Brasil. Nesse momento foi utilizado um software em C++ para a separar os registros pertencentes à Pernambuco. Em seguida, se deu necessário a obtenção das notas das questões objetivas e do questionário, que contém 50 perguntas sobre a situação socioeconômicas, dos participante em meio a todos os dados, foi utilizado o Excel, ferramenta da Microsoft, onde foi possível localizar e extrair essas informações. Agora com um estudo mais específico sobre os dados selecionados, notou-se que alguns campos estavam vazios, o que impedia aplicar as técnicas estatísticas da melhor forma, então foi realizado uma limpeza para a remoção das linha e colunas que não serviria para o estudo.

Após todos os procedimentos citados a cima, uma nova base foi formada por 389.986 registros de candidatos com suas respectivas notas e 42 questões socioeconômicas (ver Tabela 1). Posteriormente, a partir das quatro notas das provas objetivas (Ciências Humanas e suas Tecnologias, Ciências da Natureza e suas Tecnologias, Matemática e suas Tecnologias, Linguagens Códigos e suas Tecnologias), foi calculado a mediana das notas, também foi necessário a conversão dos dados socioeconômicos, pois, os mesmos, estavam representados de forma categórica com caracteres, então foram convertidos para categóricos numéricos, esses procedimentos

foram realizados utilizando as ferramentas do Excel para facilitar a posterior aplicação dos métodos estatísticos.

Tabela 1. Questões socioeconômicas selecionadas para o estudo

ID	Nº	Discriminação
1	Q001	Até que série seu pai, ou o homem responsável por você, estudou?
2	Q002	Até que série sua mãe, ou a mulher responsável por você, estudou?
3	Q003	A partir da apresentação de algumas ocupações divididas em grupos ordenados, indique o grupo que contempla a ocupação mais próxima da ocupação do seu pai ou do homem responsável por você. (Se ele não estiver trabalhando, escolha uma ocupação pensando no último trabalho dele).
4	Q004	A partir da apresentação de algumas ocupações divididas em grupos ordenados, indique o grupo que contempla a ocupação mais próxima da ocupação da sua mãe ou da mulher responsável por você. (Se ela não estiver trabalhando, escolha uma ocupação pensando no último trabalho dela).
5	Q005	Incluindo você, quantas pessoas moram atualmente em sua residência?
6	Q006	Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares.)
7	Q007	Em sua residência trabalha empregado(a) doméstico(a)?
8	Q008	Na sua residência tem banheiro?
9	Q009	Na sua residência tem quartos para dormir?
10	Q010	Na sua residência tem carro?
11	Q011	Na sua residência tem motocicleta?
12	Q012	Na sua residência tem geladeira?
13	Q013	Na sua residência tem freezer (independente ou segunda porta da geladeira)?
14	Q014	Na sua residência tem máquina de lavar roupa (o tanquinho NÃO deve ser considerado)?
15	Q015	Na sua residência tem máquina de secar roupa (independente ou em conjunto com a máquina de lavar roupa)?
16	Q016	Na sua residência tem forno micro-ondas?
17	Q017	Na sua residência tem máquina de lavar louça?
18	Q018	Na sua residência tem aspirador de pó?
19	Q019	Na sua residência tem televisão em cores?
20	Q020	Na sua residência tem aparelho de DVD?
21	Q021	Na sua residência tem TV por assinatura?
22	Q022	Na sua residência tem telefone celular?
23	Q023	Na sua residência tem telefone fixo?
24	Q024	Na sua residência tem computador?
25	Q025	Na sua residência tem acesso à Internet?
26	Q026	Você exerce ou já exerceu atividade remunerada?
27	Q034	Indique os motivos que levaram você a participar do ENEM: Testar meus conhecimentos, para aumentar as minhas chances de ingressar na Educação Superior.
28	Q035	Indique os motivos que levaram você a participar do ENEM: Ingressar na Educação

		Superior pública.
29	Q036	Indique os motivos que levaram você a participar do ENEM: Ingressar na Educação Superior privada.
30	Q037	Indique os motivos que levaram você a participar do ENEM: Conseguir uma bolsa de estudos (ProUni, outras).
31	Q038	Indique os motivos que levaram você a participar do ENEM: Participar do Programa de Financiamento Estudantil - FIES.
32	Q039	Indique os motivos que levaram você a participar do ENEM: Participar do Programa Ciências Sem Fronteiras.
33	Q040	Indique os motivos que levaram você a participar do ENEM: Aumentar a possibilidade de conseguir um emprego.
34	Q042	Em que tipo de escola você frequentou o Ensino Fundamental?
35	Q043	Marque entre as opções abaixo aquela que melhor descreve a modalidade de Ensino Fundamental que você frequentou
36	Q044	Em que turno você frequentou o Ensino Fundamental?
37	Q045	Durante o Ensino Fundamental, você abandonou os estudos e/ou foi reprovado?
38	Q046	Você já concluiu ou está concluindo o Ensino Médio?
39	Q047	Em que tipo de escola você frequentou o Ensino Médio?
40	Q048	Marque entre as opções abaixo aquela que melhor descreve a modalidade de Ensino Médio que você frequentou
41	Q049	Em que turno você frequentou o Ensino Médio?
42	Q050	Durante o Ensino Médio, você abandonou os estudos por algum tempo e/ou foi reprovado?

4.2. Aplicação Estatística

Com a base organizada foi possível buscar quais são os fatores que mais influenciam nas notas das escolas. Inicializando essa etapa, foram utilizados os dados disponíveis para realizar a correlação entre a mediana das notas e as variáveis socioeconômicas, utilizando o *software R*. Os resultados estão dispostos abaixo através de uma tabela dos coeficientes (veja a Tabela 2) e de um gráfico de dispersão (figura 2).

Tabela 2. Coeficientes de correlação

ID	Coeficientes	ID	Coeficientes	ID	Coeficientes	ID	Coeficientes
1	0,104787	12	0,042059	23	0,099021	34	0,198357
2	0,152712	13	0,124249	24	0,171239	35	-0,06264
3	0,07491	14	0,098339	25	0,117238	36	-0,15465
4	0,089444	15	0,057234	26	-0,17207	37	-0,15251
5	0,003801	16	0,085857	27	-0,00299	38	0,062658
6	0,203921	17	0,017996	28	0,046239	39	0,188524
7	0,102261	18	0,102256	29	-0,06996	40	-0,05756
8	0,166348	19	0,144762	30	-0,12498	41	-0,20476
9	0,129941	20	0,03344	31	-0,09554	42	-0,13393

10	0,160218	21	0,102498	32	0,026829		
11	0,005174	22	0,149532	33	-0,04041		

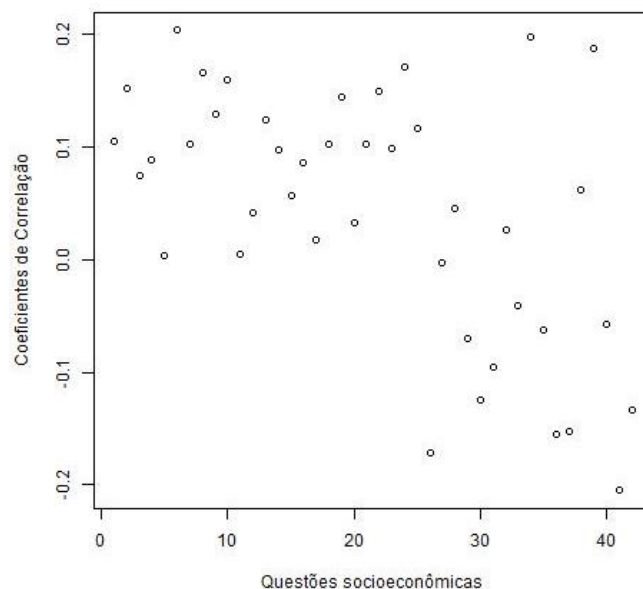


Figura 2. Gráfico de dispersão dos coeficientes de correlação

Pelos dados da correlação obtidos, nota-se que há uma correlação baixa entre os dados e uma dependência com muitos fatores. Logo com o objetivo de estudar apenas as variáveis mais influentes para o caso, foi aplicado o método de seleção *Stepwise* com direção *backward*, com a função *step()* no *software R* utilizou-se como entrada a direção e a Regressão Linear (RL). Para fazer a RL, dessa vez foi utilizado a função *lm()* do *software R*, usando como entrada a mediana das notas (variável resposta), e as variáveis socioeconômicas (variáveis explicativas).

Aplicado o método *stepwise*, das 42 variáveis que foram utilizadas, resultou na seleção de 36 variáveis, no entanto ainda são muitas variáveis. Visto que foi aplicado novamente o método *stepwise*, agora com as 36 variáveis resultantes da aplicação anterior, porém não houve eliminação de mais fatores. Em busca de reduzir as variáveis com as menores correlações, e manter apenas as mais significativas, foram escolhidas as variáveis que apresentaram a correlação maior que 15% (ver Tabela 3), mas descartadas aquelas que não obtiveram coerência lógica.

Tabela 3. Fatores socioeconômicos mais influentes

Coeficientes	Discriminação
0,203921	Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares.)
0,198357	Em que tipo de escola você frequentou o Ensino Fundamental?
0,188524	Em que tipo de escola você frequentou o Ensino Médio?
0,171239	Na sua residência tem computador?
0,152712	Até que série sua mãe, ou a mulher responsável por você, estudou?

-0,17207	Você exerce ou já exerceu atividade remunerada?
-0,20476	Em que turno você frequentou o Ensino Médio?

Na tabela acima, estão dispostas as 7 variáveis que, segundo os coeficientes de correlação, tem as maiores associações lineares com as notas das provas objetivas. Portanto, entre todos os fatores socioeconômicos analisados, esses são mais influentes.

Com a finalidade de verificar o desempenho dos métodos de mineração de dados aplicados, será possível utilizar o método de *holdout* no âmbito da simulação de Monte Carlo, este método permite, essencialmente, simular o comportamento de processos que dependem de fatores aleatórios [Sobol 1983]. No presente momento, está em análise os meios de aplicação da simulação de Monte Carlo.

5. Conclusão

Com o objetivo de Identificar os fatores determinantes no desempenho das escolas do estado de Pernambuco através dos dados do ENEM, esse projeto tem o intuito de contribuir para o desenvolvimento de políticas públicas, como também, melhoria do processo educacional no estado de Pernambuco. Estão listados abaixo os fatores socioeconômicos que mais influenciam a nota de cada participante, conseqüentemente, a nota das escolas no ENEM.

- Renda mensal da família;
- Tipo da escola frequentada no Ensino Fundamental;
- Tipo da escola frequentada no Ensino Médio;
- Existência computadores na residência;
- Nível da escolaridade da mãe, ou a mulher responsável do participante;
- Exercício de atividade remunerada;
- Turno em que era frequentado o Ensino Médio.

É importante frisar que, para a finalização do estudo, será necessário realizar a avaliação do desempenho dos métodos de *Data Mining* utilizados, através de procedimentos já citados na terceira seção desse artigo.

6. Referências

Portal do Inep. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-acessar> acessado em 01/11/2016 as 18:00 hs.

Douglas C. Montgomery. Estatística aplicada e probabilidade para engenheiros, 4ª Ed., Editora LTC, 2003.

PETER DALGAARD. (2002) Introductory Statistics with R. SpringerKnuth, D. E. (1984), The TeXbook, Addison Wesley, 15th edition.

Helene, O. (2006). “Métodos dos Mínimos Quadrados”, Editora Livraria da Física.

- Microdados. Portal do Inep Disponível em: <http://portal.inep.gov.br/microdados> acessado em 12/10/2017 às 15:00 hs.
- Enciclopédia Culturama, Definição, Conceito, Significado, o que é Correlação. Disponível em: https://educavita.blogspot.com.br/2013/08/correlacao_1792.html acessado em 12/10/2017 às 20:00 hs.
- Lori Viali, Dr. Estatística Básica, Texto V: Correlação e Regressão. Disponível em: http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_5.pdf acessado em 12/10/2017 às 21:00 hs.
- Reinartz, T.; Wirth, R.; Clinton, J.; Khabaza, T.; Hejlesen, J.; Chapaman, P.; Kerber, R. (1998). The current CRISP-DM process model for data mining. Berlin: Technical University.
- I. Sobol, "O método de Monte Carlo," *Editora Mir*, 1983.
- Devmedia. Conceitos e Técnicas sobre Data Mining. Disponível em: <http://www.devmedia.com.br/conceitos-e-tecnicas-sobre-data-mining/19342> acessado em 13/10/2017 às 14:00 hs.
- Portal Action. Seleção Backward. Disponível em: <http://www.portalaction.com.br/analise-de-regressao/2722-selecao-backward> acessado em 13/10/2017 às 14:00hs.
- Chernick, M. R. 1999. Bootstrap Methods: A Practitioner's Guide, New York: John Wiley & Sons.
- Globo. Educação, ENEM 2015: Notas por Escolas. Disponível em: <http://especiais.g1.globo.com/educacao/2016/enem-2015-notas-por-escolas/> acessado em 19/10/2017 às 20:00 hs.

Coding Dojo aplicado na monitoria como ferramenta para o aprendizado de programação integrado com o ensino a distância

Luan Augusto Xavier de Oliveira¹, Jorge Cavalcanti Barbosa Fonsêca¹

¹Universidade de Pernambuco (UPE) – Caruaru – PE – Brasil

luan.augusto.xavier.oliveira@gmail.com, jorge.fonseca@upe.br

Abstract. *This paper presents the experience of using dojo coding as a way to stimulate the practical understanding of Object Oriented Programming in Software Development in the University of Pernambuco, campus Caruaru, besides applying a Distance Learning methodology, has been produced a digital video collection that was subsequently published through the YouTube, a platform by Google, for providing the students an extra way to acquire knowledge.*

Resumo. *Este artigo apresenta a experiência de se utilizar coding dojo como forma de estimular o entendimento prático da disciplina de Programação Orientada a Objetos em Desenvolvimento de Software na Universidade de Pernambuco, campus Caruaru, além disso aplicando a metodologia de Ensino a Distância, foi realizada a produção de um acervo de vídeos digitais que subsequentemente foram disponibilizados através da plataforma YouTube, do Google, para que os estudantes tenham uma forma extra de adquirir conhecimentos.*

1. Introdução

Quando estudantes dos cursos de graduação relacionados à computação começam a aprender algoritmos, linguagens estruturadas de programação são apresentadas. Neste contexto, monitores de disciplina são importantes pois um novo universo está sendo apresentado, trazendo desafios que necessitam de prática e acompanhamento, tornando o aprendizado direto e eficaz "[...] a monitoria é importante e pode ser usada como ferramenta didático-pedagógica no processo de ensino e aprendizagem nos cursos de graduação." [de Lima e de Medeiros Simões 2016]. Com esse entendimento consolidado, um novo paradigma de programação é apresentado: o mundo orientado à objetos.

Linguagens de programação orientada à objetos são atualmente as mais utilizadas para desenvolvimento de *software* no mercado, tanto por prover um alto nível de organização e com a grande vantagem de poder reaproveitar códigos, como por permitir ao programador a capacidade de modelar características do mundo real de forma fiel. [Borges 2000] fala que o mercado de trabalho está cada vez mais interessado em trabalhar com linguagens orientadas a objetos, e utilizando elas desde do início do aprendizado fica mais evidente toda a capacidade e benefícios de utilizá-la. Sendo assim, é importante que haja um entendimento consolidado possibilitando aproveitar todas as vantagens que o paradigma oferece.

Quando o paradigma orientado a objetos é introduzido, dúvidas de sintaxe, modelagem e abstração surgem, pois a forma que um problema tem de ser pensado e analisado

já não é mais o mesmo. Assim, a prática se torna fundamental para o entendimento de novos conceitos. Mais uma vez, o papel do monitor de disciplina se torna essencial na construção do conhecimento que será a base para futuras disciplinas e profissão, [Vailati 2010] diz que a programação se aprende de forma prática, os livros e materiais não devem ser dispensados, mas deve ser levado em consideração que uma das melhores formas de se aprender programação é programando.

No mercado de trabalho, uma técnica bastante utilizada na construção e compartilhamento de conhecimento é a técnica de *coding dojo*, onde a prática colaborativa é priorizada, [Corrêa et al. 2015] afirma que realizar atividades em grupo promove o interesse dos participantes e o trabalho em equipe. Nesta, todos os envolvidos podem participar para solucionar um problema em comum, permitindo a troca de experiências e incentivando o aprendizado, desenvolvendo os envolvidos tanto socialmente quanto tecnicamente.

Este trabalho tem como objetivo apresentar uso da técnica *coding dojo* para estimular o aprendizado do paradigma orientação à objetos, além de estender o envolvimento dos discentes para além do ambiente físico da disciplina através do AcervoUPE¹, explorando ainda mais os recursos tecnológicos para melhorar a interação entre todos os atores desse sistema.

2. Trabalhos Relacionados

No trabalho de [da Silva e Endo 2016] foi apresentado o processo de produção e distribuição de material digital no AcervoUPE, assim como a sua importância na construção do aprendizado dos discentes.

Já no artigo publicado por [Carmo e Braganholo 2012] foi apresentado a experiência de se utilizar o *coding dojo* como uma ferramenta didática aplicada em disciplinas de programação.

Já [Delgado et al. 2012] mostra em seu trabalho a utilização de *coding dojo* no ensino superior e como cada disciplina influencia no problema que será trabalhado durante as sessões.

É possível observar no trabalho de [Schoeffel et al. 2016] a significância de se utilizar o *coding dojo* no aprendizado de programação orientada a objetos através de um experimento.

3. Materiais e Métodos

A disciplina de Orientação à Objetos no Desenvolvimento de Software (OODS) é ofertada apenas no segundo semestre do ano, logo, o primeiro semestre foi utilizado para preparação do conteúdo digital EAD para o AcervoUPE, que futuramente servirá aos estudantes como fonte de estudos; já durante o segundo semestre a monitoria é realizada de forma presencial e *online* para os discentes matriculados.

3.1. Sobre a disciplina

A disciplina de OODS possui uma carga horária de 60h, que são divididas em 18 aulas com 4h cada, os conteúdos são divididas de forma que os principais conceitos do

¹Canal no Youtube com conteúdo específicos de disciplinas

paradigma sejam apresentados, além disso, por ser a primeira disciplina onde há uma completa imersão no paradigma orientado a objetos é necessário uma ênfase em conceitos mais básicos que serão aprofundados em outras disciplinas no decorrer do curso.

Em versões anteriores da disciplina já foi constatado que uma das maiores dificuldades dos estudantes é que os mesmos não conseguem executar o que foi visto em aula, dessa forma, diversos exercícios são propostos na tentativa de envolver teoria e prática permitindo que os discentes tenham um melhor aproveitamento, tornando as aulas mais dinâmicas servindo também como meio de consolidar os novos conceitos apresentados.

A linguagem de programação escolhida para apresentação dos conceitos é Java, por ter suporte nativo a OO e ser amplamente utilizada no mercado, além disso é uma linguagem de fácil utilização por possuir um alto nível de abstração, no entanto, o entendimento da mesma é mais uma dificuldade para os discentes que estão começando a entender características que até então eram desconhecidas.

3.2. *Coding dojo*

Ainda na tentativa de solucionar as dificuldades apresentadas no tópico 4.1 as sessões de *coding dojo*, ou simplesmente *dojo*, nas monitorias permite que os estudantes possam praticar de forma colaborativa, é uma técnica que vem sendo utilizada por diversas empresas de desenvolvimento de *software* a fim de que a equipe possa trocar técnicas e conhecimentos de forma prática.

O funcionamento das sessões é simples, dois participantes são selecionados como piloto e copiloto, o piloto que vai ter o controle do que está acontecendo no momento, já o copiloto tem a função de auxiliar em caso de dúvidas, os demais participantes da sessão deverão ficar em silêncio apenas observando, se necessário, a dupla irá pedir ajuda e só assim que os outros participantes poderão colaborar, promovendo a troca de conhecimentos entre os membros.

Cada dupla tem um tempo predeterminado, a ideia é que todos possam produzir, logo, tem haver o revesamento de times, ao final do tempo estipulado o participante que estava com o papel piloto irá para a plateia e o copiloto irá assumir a posição de piloto, da mesma forma outro participante da sessão irá assumir a posição de copiloto e assim por diante até o final da sessão.

A princípio foi proposto que as sessões de *dojo* fossem gravadas e postadas no AcervoUPE juntamente com os materiais, no entanto, por limitações do ambiente e de *hardware* (equipamentos para gravação e captação do áudio dos participantes) não foi possível prosseguir com as gravações.

De forma geral, o planejamento de desenvolvimento da sessão ocorre de forma semelhante que podem ser resumidas em 5 etapas:

- **Escolha do problema:** dias antes da sessão o professor e o monitor conversam sobre o último conteúdo visto em sala de aula pelos estudantes, então, o monitor fica responsável pela escolha do problema que antes da monitoria ainda deve ser repassado para o professor, fazendo modificações se necessário.
- **Propondo o problema:** no início da sessão de *coding dojo* o problema é descrito para os discentes de forma que todos tenham o entendimento claro do que vai ser trabalhado, possíveis dúvidas devem ser tiradas para que não atrapalhe no

andamento da sessão, pequenas modificações no problema podem ser realizadas, é o momento que deve haver a discussão entre todos os participantes, essa etapa pode levar de 10 à 20 minutos.

- **Modelando o problema:** já utilizando o reveasamento de duplas, compostas pelo piloto e copiloto, os estudantes vão ao quadro para modelar o problema da melhor forma possível, geralmente 30 à 40 minutos já são suficientes, é importante que todos estejam de acordo e que o entendimento do problema esteja realmente consolidado.
- **Codificação:** para a evolução da sessão é necessário mais uma vez que haja o reveasamento entre as duplas, só que agora, a ferramenta será um computador que será compartilhado por todos da sessão, dependendo do problema o tempo pode variar, mas essa etapa leva aproximadamente 1 hora e no máximo 1 hora e meia.
- **Discussão:** ao final, os participantes conversam entre si, com a presença do monitor relatando o que foi aprendido e sugestões para a próxima sessão, esse momento leva cerca de 5 a 10 minutos.

Durante a sessão, o piloto e o copiloto são sempre incentivados a compartilhar a conversa e o raciocínio com todos os participantes da sessão, permitindo assim que todos tenham conhecimento constante do que está sendo elaborado.

O código produzido em cada sessão é compartilhado no grupo da disciplina de forma que os discentes que não puderam participar possam ter acesso ao conteúdo produzido, permitindo que eles sejam capazes de estudar, otimizar e colaborar com a monitoria de forma *online*.

A técnica do *dojo* pode ser facilmente adaptada para o contexto a ser trabalhado, no trabalho de [Sato et al. 2008] é apresentado a experiência de se utilizar *coding dojo* em uma Universidade, apresentando também a metodologia que o mesmo utilizou para desenvolver as sessões.

3.3. AcervoUPE e EAD

O AcervoUPE é um canal dentro da plataforma *Youtube*, do *Google*, que já foi utilizada em outros projetos de monitoria, o mesmo já conta com conteúdos de programação estruturada em C, redes de computadores, inglês técnico e agora mais uma vez em programação, sendo orientada a objetos.

O material para o Acervo foi produzido seguindo o cronograma apresentado no projeto de monitoria, mas devido a problemas que foram surgindo, atrasos ocasionais aconteceram, no entanto, a disciplina de OODS é ofertada apenas uma vez por ano pela UPE Caruaru, logo, não atrapalhou a utilização do material pelos discentes, além disso alguns vídeos foram adaptados e divididos para evitar a postagem de materiais extensos, as principais dificuldades encontradas foram:

- **Falta de experiência:** o material utilizado em cursos EAD não é o mesmo utilizado em sala de aula, o mesmo tem de ser elaborado de forma que leve ao ouvinte o claro entendimento do que é apresentando, pela inexperiência, após a gravação da aula era constatado que o material não tinha a qualidade esperada, logo era necessário a regravação e em alguns casos a elaboração de um novo roteiro.
- **Problemas com áudio:** interferências externas, como ruídos e barulhos no ambiente, além de problemas como estática e a qualidade do equipamento utilizado

Tabela 1. Cronograma de vídeos

Aula 01	Vídeo-aula Introdução a OO
Aulas 02 e 03	Vídeo-aula Classes e Objetos
Aula 4	Vídeo-aula Tipos de Referência, String e Array
Aula 5	Vídeo-aula Relacionamento entre Objetos
Aulas 06 e 07	Vídeo-aula Herança, Polimorfismo e ligação dinâmica
Aula 08 e 09	Vídeo-aula Classe Abstrata e Interface
Aula 10	Vídeo-aula Pacotes e Exceções
Aula 11	Vídeo-aula Arquitetura em Camadas
Aula 12	Vídeo-aula Interface Gráfica

atrapalharam de forma impactante com o cronograma, mesmo quando o conteúdo era apresentado de forma satisfatória, o áudio foi um dos fatores para decidir se o material seria publicado ou se era necessário uma nova gravação para corrigir as falhas, a utilização de uma sala da própria universidade como estúdio colaborou com o avanço do cronograma.

- **Teoria e prática:** o conteúdo, quando não apresentado da forma ideal, se torna cansativo, principalmente nas aulas onde prevalece a teoria, o que foi outro fator levado em consideração nos vídeos; se a aula estava cansativa iria diminuir o interesse dos discentes, logo era necessário realizar alterações na apresentação e no roteiro antes de publicar o material.

Para garantir que o material iria servir como reforço ao apresentado em sala de aula, os vídeos são liberados de acordo com o andamento do conteúdo, dessa forma os vídeos vão servir como um material complementar ao que foi visto em sala de aula.

Sendo assim, o atraso na liberação dos vídeos foi proposital como forma de minimizar a evasão, tendo em vista que os discentes ao assistir o conteúdo *online* não crie a expectativa de que será exatamente igual ao que será visto em sala e deixe de participar das mesmas por acreditar já saber o suficiente do conteúdo, o material do acervo não substitui a aula presencial.

3.4. Sobre a Monitoria

A monitoria acontece uma vez na semana de forma presencial, mas se estende durante toda a semana de forma *online*, uma vez que as dúvidas serão postadas em um grupo preparado para disciplina em uma rede social, todos os discentes que estão cursando OODS e fazem parte desse grupo podem interagir e também acompanhar as dúvidas dos outros estudantes, participando e colaborando se achar necessário.

Durante aulas de monitoria a técnica do *coding dojo* foi aplicada de forma que os novos conceitos que foram apresentados fiquem em contraste com as técnicas de programação anteriormente aprendidas pelos discentes.

Na modalidade presencial, a monitoria acontece no laboratório da própria universidade, como já é um ambiente familiar para os estudantes, a timidez não é um fator que atrapalha a interação dos mesmos, no entanto, na modalidade *online*, há um certo receio

por parte dos estudantes em publicar suas dúvidas no grupo de forma que todos possam participar, as dúvidas são comumente enviadas como mensagem privada para o monitor.

3.4.1. Primeira aula presencial: listas encadeadas

Na primeira aula de monitoria, compareceram apenas 6 estudantes, os mesmos já haviam visto em sala de aula sobre classes e objetos, então o problema proposto precisava envolver o assunto abordado e também uma forma de evidenciar as diferenças entre as linguagens estruturadas que eles já estavam acostumados com a linguagem proposta no curso, que é Java.

Como listas encadeadas é um assunto abordado em disciplinas anteriores a OODS, foi proposto o problema de implementar essas listas, as quais eles já estavam familiarizados, mas dessa vez em Java.

Mesmo utilizando a técnica de *coding dojo*, como foi a primeira sessão e também o primeiro contato com a linguagem de programação, não houve muita limitação em relação ao diálogo entre piloto, copiloto e os demais participantes da sessão, em alguns momentos houve a necessidade da intervenção do monitor para que fosse possível dar continuidade a sessão; ao final da sessão o problema proposto foi solucionado sem grandes dificuldades.

3.5. Segunda aula presencial: *player* de músicas

Dando continuidade as sessões de *coding dojo*, mais uma vez o problema proposto tem de entrar de forma que os discentes percebam a diferença entre os paradigmas, na última aula foi apresentado os conceitos de herança e polimorfismo, dessa forma, foi definido que o problema a ser trabalhado seria um CRUD (*create-retrieve-update-delete*, mas o tema foi decidido em conjunto com os participantes da sessão, um *player* de músicas.

Mais uma vez essa sessão contou com a presença de 6 estudantes, no entanto apenas 4 haviam participado da primeira sessão, dessa vez os participantes tiveram problemas em resolver o problema proposto, precisando que da intervenção do monitor diversas vezes, os motivos observados para essa dificuldade foram:

- **Dificuldades com a linguagem de programação:** ficou evidente que a maior dificuldade durante a sessão não é o problema proposta em si, mas sim a falta de conhecimento na plataforma que está sendo utilizada;
- **Falta de atenção ao código escrito por outros participantes:** durante o revezamento, nem todos os participantes estavam totalmente engajados com o andamento da sessão, o que trouxe dificuldades para entendimento e continuação quando eram escalados como piloto;
- **Falta de entendimento do problema:** em diversos momentos, os participantes estavam fugindo do problema proposto e da modelagem definida no início da sessão.

Ao final, foi discutido sobre as dificuldades encontradas para andamento da sessão, as opiniões dos participantes foi encaminhada para o professor da disciplina, o que levou a mudanças na terceira aula presencial.

3.6. Terceira aula presencial: gerenciamento de bibliotecas

Pelas dificuldades encontradas na segunda aula de monitoria, essa ocorreu da forma mais tradicional, a técnica de *coding dojo* não foi aplicada, mas uma aula onde dúvidas foram tiradas havendo também um momento onde o monitor desenvolveu juntamente com os estudantes um sistema para gerenciar bibliotecas.

Além disso, uma atividade similar foi aplicada para aproveitar parte do tempo da monitoria e também para que os discentes pudessem praticar os conceitos que foram apresentados em casa, incentivando assim que dúvidas sejam levadas para sala de aula e também para as próximas monitorias promovendo uma melhor interação entre alunos-professor-monitor.

Além disso, mesmo havendo sessões de *coding dojo*, mais atividades serão propostas aos final das sessões e divulgadas no grupo da disciplina, para que tanto os participantes quanto os demais discentes da disciplina possam praticar de forma autônoma, consequentemente, incentivando a monitoria *online*, já que não há o monitor de forma presencial, as redes sociais servirá de meio para solução das dúvidas.

3.7. Aulas presenciais subsequentes

As próximas aulas presenciais de monitoria, continuará sendo utilizada a técnica de *coding dojo*, as discussões e dificuldades encontradas nas sessões anteriores servirá de experiência para a melhoria das próximas, além disso, espera-se que as dificuldades em relação a linguagem de programação diminua com o passar do tempo, onde os participantes irão se familiarizar com a mesma.

No entanto, a utilização ou não de sessões de programação colaborativa irá depender da quantidade de discentes, um número pequeno de participantes torna a sessão cansativa além de perder o foco do conhecimento ser construído através da cooperação de um grupo.

Além disso, o foco das sessões não é a solução do problema em si, mas sim a troca de conhecimentos entre os participantes, onde os erros e dificuldades servirão de alicerce para construção dos conhecimentos que serão adquiridos.

3.8. Atividades propostas para prática individual

Após cada sessão de *coding dojo* os discentes recebem uma sugestão de atividade envolvendo os conceitos abordados na tentativa de estimular a prática em outros momentos fora do horário estipulado para a monitoria, os mesmos podem utilizar da modalidade *online* para solucionar possíveis dúvidas em relação a atividade, além disso, a atividade não é de carácter obrigatório, ou seja, é uma forma do aluno praticar sem maiores cobranças.

Quando os alunos estão todos envolvidos na sessão de *coding dojo*, dúvidas podem não ficar tão evidentes, pois devido o trabalho em conjunto detalhes do código podem passar despercebidos, logo, é importante que os mesmos tentem desenvolver problemas de forma individual para que o entendimento fique consolidado.

Além disso, as atividades não são limitadas apenas para os participantes das sessões, a sugestão é publicada no grupo da disciplina, dessa forma, todos ficam cientes da proposta a ser trabalhada.

4. Resultados

Durante a semana, é possível medir e analisar o aproveitamento dos discentes através da interação dos mesmos com o monitor na tentativa de esclarecer dúvidas, evidenciando o nível de interesse por parte dos discentes com as atividades propostas, além disso os participantes da monitoria na modalidade *online* são os mesmos que participam de forma presencial, evidenciando assim, que os participantes das sessões de *coding dojo* tem um maior incentivo na resolução de atividades propostas.

As métricas do material publicado no canal AcervoUPE mostram que os mesmos não possuem quantidades significativas de visualizações, em uma enquete realizada no grupo da disciplina que possui um total de 18 participantes, 5 responderam que assistem as aulas regularmente, 4 responderam que assistiram algumas aulas e 9 não responderam a pesquisa.

A técnica de *coding dojo* mostrou ser eficaz para a construção de conhecimento de forma colaborativa, durante as sessões também é possível identificar competitividade entre os participantes onde os mesmos se esforçam para não deixar parte do seu código para que os demais participantes concluam.

Vale ressaltar que o presente artigo foi elaborado durante o período letivo da disciplina, logo, não é possível tirar resultados e conclusões definitivas, sendo assim, as informações inferidas tem carácter parcial.

5. Conclusão

As visualizações do material da *playlist* de Orientação a Objetos produzido para o AcervoUPE tem um baixo número de visualizações, logo é possível concluir que não houve um incentivo para que os discentes assistissem ao conteúdo ou as aulas presenciais e monitorias já são suficientes para o entendimento do conteúdo, logo não houve interesse por parte dos discentes em assistir.

Mesmo com as dificuldades apresentadas, é viável a utilização da técnica de *coding dojo* nas monitorias, desde que haja algumas modificações na técnica original, invés de restringir o diálogo apenas para o piloto e copiloto, permitir que todos os participantes possam discutir livremente ajudando quando achar necessário.

Referências

- Borges, M. A. (2000). Avaliação de uma metodologia alternativa para a aprendizagem de programação. In *VIII Workshop de Educação em Computação-WEI*.
- Carmo, D. H. e Braganholo, V. (2012). Um estudo sobre o uso didático de dojos de programação. In *Workshop de Educação em Computação (WEI). Sociedade Brasileira de Computação*.
- Corrêa, A. G. D., de Almeida Pacheco, B., et al. (2015). Coding dojo como metodologia de ensino: A visão dos educadores. *Revista Educação & Tecnologia*.
- da Silva, L. G. F. e Endo, P. T. (2016). Acervoupe: Produção de material online para auxílio e acompanhamento do componente curricular redes de computadores i. *JCE*, pages 1–6.

de Lima, J. E. C. e de Medeiros Simões, A. S. (2016). A monitoria como ferramenta didático-pedagógica no processo de ensino/aprendizagem nos cursos de graduação. pages 1–8.

Delgado, C., de Toledo, R., e Braganholo, V. (2012). Uso de dojos no ensino superior de computação.

Sato, D. T., Corbucci, H., e Bravo, M. V. (2008). Coding dojo: An environment for learning and sharing agile practices. In *Agile 2008 Conference*, pages 459–464.

Schoeffel, P., Rosa, D. F., e Waslawick, R. S. (2016). Um experimento do uso de coding dojo na aprendizagem de programação orientada a objetos. *iSys-Revista Brasileira de Sistemas de Informação*, 9(2).

Vailati, T. (2010). Aprenda mais com coding dojo. pages 1–2.

O uso do KNoT como uma plataforma tradicional de IoT

João Neto¹, Lucas Feijó¹, Jorge Fonseca¹, Fernando Carvalho¹

¹Sistemas de Informação - Universidade de Pernambuco (UPE)
Caruaru - PE - Brasil

{joao.alexandre, jorge.fonseca}@upe.br

{lucasfjportela, fernandocarvalhobr}@gmail.com

Abstract. *The Internet of Things (IoT) has emerged with the goal of forming a global network of connected intelligent objects. For this to be possible, several devices with totally different capabilities and requirements need to be connected. Due to this heterogeneity of technologies, and also to solve problems related to data security, analysis and processing, IoT platforms have been developed. However, with the constant development of platforms, another problem of heterogeneity of solutions is verified by the market and the academy. To solve this, the meta-platform KNoT was developed by CESAR in order to integrate all these solutions. However, KNoT in its current stage integrates only one IoT platform. This work aims to develop a Smart House application to validate the operation of KNoT in a scenario of integration with a single platform.*

Resumo. *A Internet das Coisas (Internet of Things - IoT) surgiu com objetivo de formar uma rede global de objetos inteligentes conectados. Para que isso seja possível, diversos dispositivos com capacidades e requisitos totalmente diferentes precisam estar conectados. Devido a essa heterogeneidade de tecnologias, e também para resolver problemas relacionados a segurança, análise e processamento dos dados, foram desenvolvidas plataformas de IoT. No entanto, com o constante desenvolvimento de plataformas, outro problema de heterogeneidade de soluções é constatado pelo mercado e pela academia. Para resolver isso, a metaplataforma KNoT foi desenvolvida pelo CESAR com objetivo de integrar todas essas soluções. No entanto, o KNoT em seu estágio atual, integra apenas uma plataforma de IoT. Este trabalho visa desenvolver uma aplicação de Casa Inteligente para validar o funcionamento do KNoT em um cenário de integração com uma única plataforma.*

1. Introdução

Nos últimos anos, diversas soluções de Internet das Coisas (Internet of Things - IoT) têm surgido e motivado o desenvolvimento de aplicações inovadoras em vários cenários. Isso tem acontecido a partir da concretização da ideia inicial de Computação Ubíqua, proposta por [Weiser 1991], que diz respeito ao nível de integração de equipamentos com capacidade de processamento em nosso cotidiano, chegando a serem imperceptíveis. A IoT vai além, e integra capacidade de comunicação com a Internet a esses dispositivos já autônomos, promovendo soluções mais inovadoras e de múltiplos domínios.

Um dos principais desafios nesse contexto, está em como integrar e oferecer mecanismos de segurança e processamento de dados a dispositivos com requisitos tecnológicos (*hardware* e protocolos de comunicação) totalmente diferentes. Para resolver isso, foram desenvolvidas as plataformas, que possuem como objetivos principais resolver esses problemas e facilitar o desenvolvimento das aplicações. O mercado, atualmente, conta com uma grande diversidade de plataformas de IoT [Mineraud et al. 2016], específicas a cada contexto.

De acordo com o instituto de pesquisa McKinsey [Manyika 2015], 40% do potencial da IoT, em termos de valor gerado, pode ser viabilizado apenas com um cenário de interoperabilidade. Sendo assim, plataformas heterogêneas - de múltiplos domínios - deveriam se integrar com o objetivo de gerar mais valor, uma vez que múltiplas fontes de dados podem produzir melhores *insights*. A partir disso, soluções como a metaplataforma KNoT (KNoT *Network of Things*), desenvolvida pelo CESAR, pretende integrar múltiplas plataformas e resolver esse problema em potencial.

No entanto, em sua primeira versão, o KNoT possui integração com apenas uma plataforma de IoT. A proposta desse trabalho é verificar a capacidade de funcionamento do KNoT como uma plataforma típica de IoT, ou seja, capaz de viabilizar o desenvolvimento de aplicações voltadas para um único cenário. Para isso, implementamos uma aplicação simples voltada para o cenário de Casas Inteligentes. O trabalho está organizado da seguinte forma. Na seção dois é feita uma revisão de literatura sobre os temas em questão. Nas seções 3 e 4 abordamos a arquitetura do KNoT e a aplicação para validar nossa proposta. Na seção 5 nós demonstramos a aplicação em funcionamento. Por fim, na seção 6 concluímos e mencionamos possíveis trabalhos futuros.

2. Revisão de Literatura

2.1. Internet das Coisas

Nas últimas décadas, presenciamos o processo de miniaturização de dispositivos eletrônicos com capacidade de sensoramento, atuação e comunicação em pequenas distâncias. Esses dispositivos se integraram naturalmente ao cotidiano das pessoas, possibilitando o surgimento de um paradigma chamado Computação Ubíqua, que representa a percepção nula dessa integração, tendo como ponto de vista as pessoas [Weiser et al. 1999].

Foi nesse contexto que objetos do cotidiano passaram a ser, além de autônomos, endereçados unicamente através da Internet. Esse fenômeno ganhou forma e iniciou um novo paradigma, chamado Internet das Coisas. O termo foi cunhado por Kevin Ashton, co-fundador e ex diretor executivo do Auto-ID Center, enquanto estava apresentando uma

palestra na Procter & Gamble (P&G), sobre a utilização de RFID na logística da empresa. De acordo com [Atzori et al. 2015], fez uma classificação da IoT em três principais orientações: *thing*, semântica e internet.

- *Thing*: essa perspectiva enxerga a IoT como sendo composta de meros objetos identificados por *tags* RFID (*Radio-Frequency IDentification*).
- *Internet*: evidencia a Internet e as tecnologias de rede desenvolvidas, como sendo os principais elementos em uma possível definição da IoT.
- *Semântica*: enfatiza os desafios com relação à representação, armazenamento, interconexão, pesquisa e organização da grande quantidade de informação gerada pela IoT.

Ainda segundo este autor, a convergência entre essas três visões irá representar a IoT em sua totalidade.

2.2. Plataformas de IoT

As plataformas de IoT foram desenvolvidas com o objetivo de integrar objetos com tecnologias e protocolos heterogêneos, além de oferecer mecanismos que resolvem problemas relacionados a segurança, visualização e processamento dos dados. Há uma multiplicidade de plataformas disponíveis no mercado, como pode ser visto em [Mineraud et al. 2016]. Para o presente estudo, iremos analisar e comparar três soluções (ThingSpeak, IBM Bluemix e KNoT), escolhidas por conveniência, onde duas são plataformas típicas e bem conhecidas e outra surge com um propósito maior (integrar as demais plataformas), mas poderá ser abordada, em nosso contexto, como uma plataforma equivalente às demais.

A *ThingSpeak* é uma plataforma *open-source*, que tem o servidor como elemento principal de sua arquitetura, usado para receber e armazenar dados. Ele provê suporte a troca de dados em múltiplos formatos (JSON, XML e CSV), além de oferecer recursos de visualização dos dados, possibilitando ainda a criação de *widgets* em linguagens como JavaScript, para visualização personalizada dos dados [Mineraud et al. 2016].

A IBM Watson IoT é uma plataforma de nuvem proprietária e com pegada forte no oferecimento de serviços poderosos para visualização e processamento de dados. Sua arquitetura é flexível, ou seja, o desenvolvedor pode tanto conectar seus dispositivos diretamente na *Cloud*, quanto conectar seus dispositivos por meio de um *Gateway*, seguindo o modelo de arquitetura mostrado na Figura 1. Nessa plataforma, os dados podem ser trafegados por meio de um protocolo leve como o MQTT, ou utilizando um protocolo tradicional como o HTTP. As aplicações podem coletar os dados ou realizar operações nos objetos por meio de uma API REST. A IBM Watson fornece também uma interface web para gerenciamento dos dispositivos e acompanhamento dos dados em tempo real.

Em face a grande quantidade de plataformas disponíveis [Mineraud et al. 2016], e as que vão continuar sendo desenvolvidas de acordo com a [Gartner 2017], o KNoT (KNoT *Network of Things*) foi desenvolvido pelo CESAR¹ com objetivo de integrá-las em prol da interoperabilidade entre plataformas de IoT. Ela é *open-source*, multi-protocolo e desenvolvida para *hardware* de baixo custo. Sua arquitetura também é baseada no modelo mostrado na Figura 1.

¹<http://cesar.org.br>

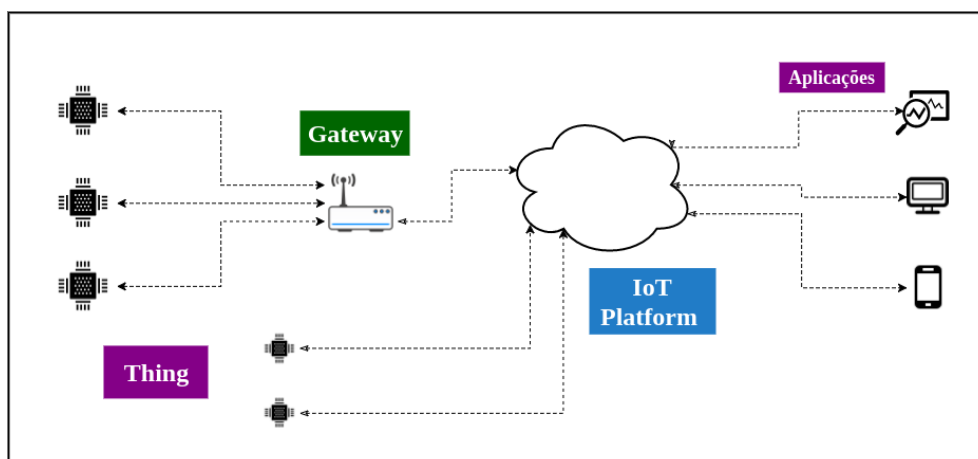


Figura 1. Arquitetura Geral de IoT

A *ThingSpeak* tem suas vantagens em termos de facilidade de uso, mas não oferece uma arquitetura tão flexível. Os dispositivos precisam enviar os dados diretamente para o servidor, através do protocolo HTTP. Por outro lado, o IBM Watson IoT, além de suportar a integração de gateway, permite o uso do protocolo MQTT (mais eficiente em IoT) e fornece poderosos recursos de visualização e processamento de dados. Por fim, o KNoT reúne todas as vantagens e ainda provê mais flexibilidade, uma vez que múltiplas plataformas podem ser integradas, munindo os desenvolvedores com diversos recursos de *cloud*. Diante disso, para a presente pesquisa, o KNoT foi escolhido.

3. Arquitetura do KNoT

A arquitetura da metaplataforma KNoT apresenta três componentes principais, como pode ser visto na Figura 2. O componente *Device* representa o objeto inteligente, responsável pelo sensoriamento e/ou atuação sob o ambiente. Esse *Device* é conectado a Internet através do componente *Gateway*, que entre outras coisas também armazena e processa dados recebidos do *Device* antes mesmo de enviar para a *Cloud*. Este componente, por último, representa a plataforma de nuvem acoplada à metaplataforma.

3.1. Device

Este primeiro tem a função de representar o dispositivo inteligente e seus requisitos. De acordo com a arquitetura, será composto por um microcontrolador, alimentado por uma fonte de energia, e módulos de rádio acoplados para possibilitar a comunicação entre o dispositivo e outros componentes. A nível de *software*, será controlado pelo micro sistema operacional do KNoT, através de uma camada de abstração de hardware (Hardware Abstraction Layer - HAL). Os dados gerados pelo *Device* serão enviados via protocolo específico do KNoT para o *Gateway*, explicado na próxima subseção.

3.2. Gateway

O *Gateway* tem papel fundamental em três aspectos: endereçamento dos dispositivos na Internet, tradução e processamento dos dados e integração com as plataformas de *Cloud*. A princípio, este componente atribuirá um endereço IP a cada dispositivo que não possua conexão direta com a Internet, uma vez que ambos estejam pareados. Após isso, os dados

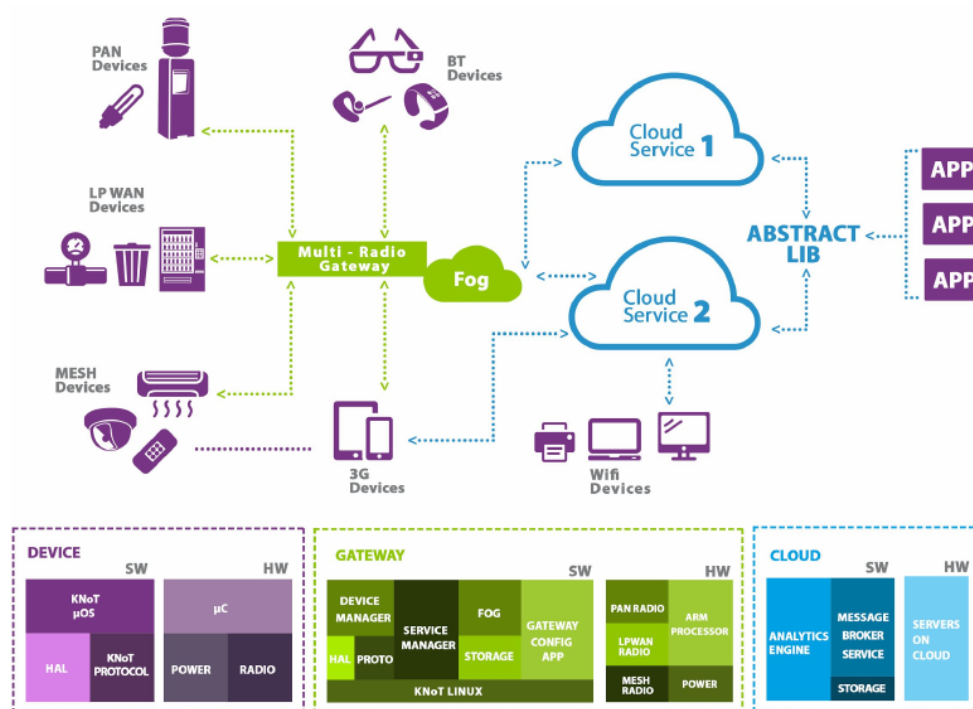


Figura 2. Arquitetura do KNoT

recebidos do *Device*, através do protocolo KNoT, serão traduzidos para formatos semi-estruturados, como JSON ou XML, para facilitar a leitura pelos serviços de *Cloud*. Os dados podem não precisar serem enviados diretamente para processamento na plataforma integrada, uma vez que uma instância reduzida da mesma, denominada *Fog*, estará sendo executada no *Gateway*.

3.3. Cloud

Por último, este componente será responsável por armazenar e processar os dados coletados, bem como dispor recursos para que as aplicações realizem operações nos dispositivos pareados. As especificações dos itens relacionados a este componente variam de acordo com a plataforma integrada. No entanto, em alto nível, deverá conter além de um banco de dados interno, uma *engine* para visualização, análise e processamento dos dados. Estes dados coletados podem ser utilizados por aplicações, através de requisições à API (Interface de Programação) disponibilizada pela *Cloud* e subscrições no *broker* de mensagens.

4. Aplicação

A presente aplicação visa demonstrar a utilização do KNoT para o desenvolvimento de aplicações IoT voltadas para o contexto de Casas Inteligentes. Para isso, realizamos o controle inteligente de um ventilador de acordo com a presença de uma pessoa no ambiente. A seguir descrevemos todas as tecnologias utilizadas para viabilizar a solução, de acordo com cada componente da arquitetura do KNoT.

4.1. Device

A nível de dispositivo, utilizamos a plataforma de *hardware* Arduino, uma vez que atende todos os requisitos do KNoT em termos de processamento (microcontrolador) e

alimentação, além da facilidade por ser um placa de prototipação barata e de fácil acesso. A integração com o KNoT é feita através da biblioteca KNoT Thing. Já em termos de comunicação, será utilizado o módulo wireless NRF24L01, voltado para comunicação a curtas distâncias, PAN (Personal Area Network) e suportado pela plataforma KNoT.

O sensoriamento do ambiente será realizado através do sensor de presença PIR. Os dados obtidos pelo sensor serão encaminhados, através do protocolo KNoT, para o *Gateway*.

As tecnologias são mostradas na Figura 3.

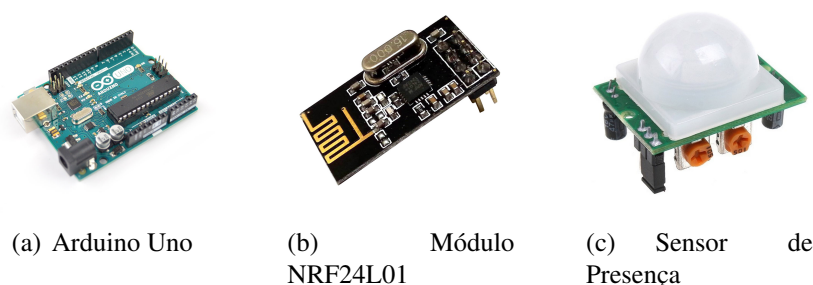


Figura 3. Tecnologias utilizadas

4.2. Gateway

O dispositivo de *hardware* utilizado como *Gateway* será o Raspberry Pi 3 (mostrado na Figura 4), suportado pelo KNoT e dotado de boa capacidade de processamento com processador Quad-Core e memória RAM de 1GB. Neste componente, também foi utilizado o módulo de comunicação NRF24L01 para viabilizar a comunicação com o *Device*. Todos os serviços correspondentes ao KNoT foram executados em cima do KNoT Linux, sistema operacional compilado especialmente para ser pouco custoso e atender todos os requisitos da arquitetura.

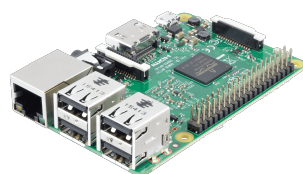


Figura 4. Raspberry Pi 3

4.3. Plataforma de IoT

A plataforma integrada é a Meshblu, sendo a única suportada pelo KNoT nesta versão. Ela é *open-source* e desenvolvida especialmente para conectar objetos inteligentes, provendo facilidades para o processamento e a análise dos dados. Para interagir com dispositivos associados a plataforma, basta saber como manipular as mensagens através de APIs bem definidas. Além disso, ela é multi-protocolo, podemos realizar operações através de HTTP ou *websocket*, por exemplo.

4.4. Aplicação Final

Fisicamente, a aplicação pode ser vista na Figura 5.

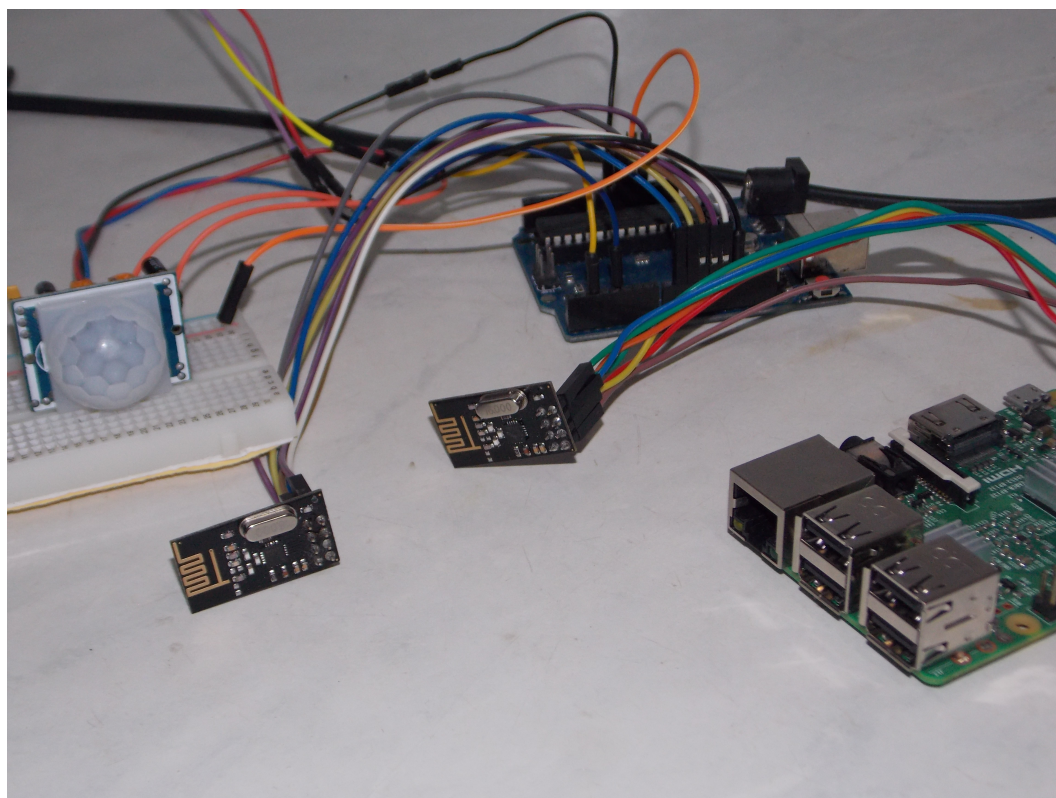


Figura 5. Aplicação Final

5. Metodologia e Resultados

Para validar a aplicação, realizamos um estudo de caso em um contexto de controle de dispositivos em Casas Inteligentes, seguindo a metodologia proposta por [Runeson and Höst 2009]. Assim, procuramos responder nossa pergunta de pesquisa: **a metaplataforma KNoT possui capacidade para funcionar como uma plataforma típica de IoT?**

Os dados foram coletados de maneira planejada, a partir de cada interação com o KNoT. Em resultados, analisamos os dados coletados a fim de chegar a uma resposta para a pergunta de pesquisa.

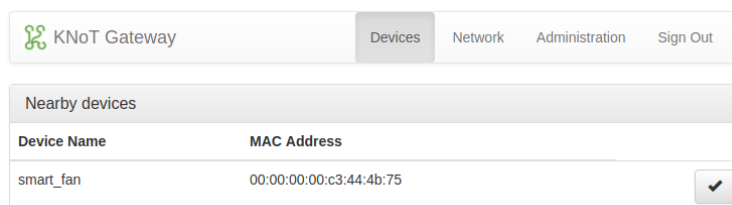
Mostraremos primeiro a integração do dispositivo com o *Gateway* e em seguida a realização de operações no objeto conectado através da plataforma de *Cloud* integrada. As operações foram realizadas a partir de uma API disponibilizada pela Meshblu e logo após executamos uma aplicação responsável por ligar o ventilador de acordo com a presença ou não de pessoas no ambiente.

5.1. Configuração do Dispositivo

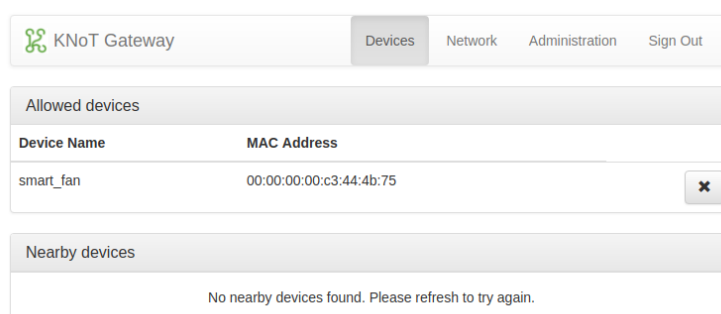
No dispositivo (Arduino), importamos a biblioteca KNoT Thing e integramos o módulo de rádio e o sensor de presença já mencionados. Para que ele realizar operações requisitadas pela *Cloud*, implementamos três funções: duas que fazem a leitura e atuação no ventilador e uma para a leitura dos dados obtidos pelo sensor de presença. Todas as funções foram registradas e associadas através de funções da biblioteca. Este dispositivo possui o nome **smart_fan**.

Uma vez que o dispositivo está em operação como um KNoT Thing, enviará constantemente sinais de presença para que o *Gateway* o reconheça e consiga estabelecer uma conexão.

5.2. Pareamento



(a) Dispositivos Próximos



(b) Dispositivo Pareados

Figura 6. Interface Web do Gateway

Na tela de configuração do *Gateway* podemos ver quais dispositivos estão sendo detectados no momento. De acordo com a Figura 6, o dispositivo *smart_fan* está sendo detectado. Ao realizar o pareamento do dispositivo, vemos que ele aparece na seção de dispositivos permitidos. No log gerado pelo *Gateway*, após o pareamento, obtemos dados importantes de autenticação para que realizemos operações no dispositivo através da *Cloud*, conforme Figura 7.

```
Oct 15 02:48:44 knot.daemon.info nrfd[6630]: Thing sending presence. MAC = 00:00:00:00:c3:44:4b:75 Name = thingname
Oct 15 02:48:44 knot.daemon.info knotd[17822]: Connecting to knot-test.cesar.org.br:3000...
Oct 15 02:48:45 knot.daemon.info knotd[17822]: LWS_CALLBACK_CLIENT_ESTABLISHED
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS JSON_RX 4 =
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS JSON_RX 4 = [{"identify":}]
Oct 15 02:48:45 knot.daemon.info knotd[17822]: node:0x1a2c728 proto:0x1a2c780
Oct 15 02:48:45 knot.daemon.info knotd[17822]: KNOT_OP: 0x14 LEN: 4c
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS TX JSON [ "identity", { "uuid": "64ed82d8-5200-431b-a3f2-ec937290000", "token": "a98738ac5719d7fca2797d02a8fb89558d49453" } ]
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS TX122 bytes
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS JSON_RX 4 = [{"ready",{"api":"connect","status":201,"socketid":"lvTDTDxgGjdajJsAABB","uuid":"64ed82d8-5200-431b-a3f2-ec937290000","cloud":"53bd27de-299c-4142-9fbb-99fcd2b4cfff","token":"a98738ac5719d7fca2797d02a8fb89558d49453"}]}]
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS JSON_TX [ "device", { "uuid": "64ed82d8-5200-431b-a3f2-ec937290000" } ]
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS TX67 bytes
Oct 15 02:48:45 knot.daemon.info knotd[17822]: WS JSON_RX 4 = [{"config",{"uuid":"64ed82d8-5200-431b-a3f2-ec937290000","online":true,"type":"KNOTDevice","name":"KNoTThing","owner":"71061601-4729-4348-8130-216a24c90000","socketid":"lvTDTDxgGjdajJsAABB","ipAddress":"","...}]
```

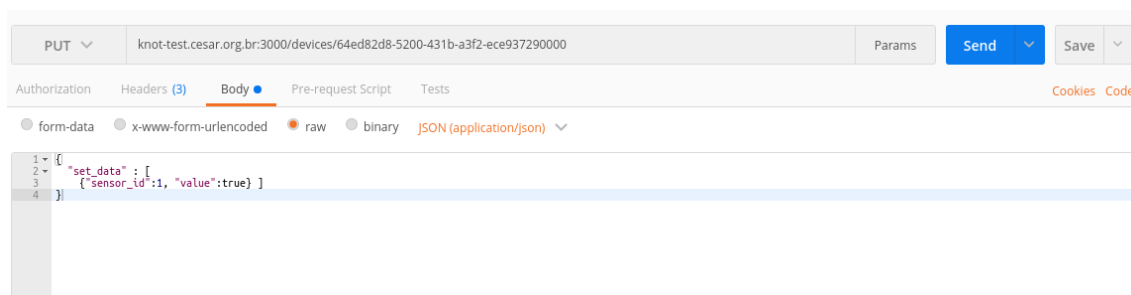
Figura 7. Log do Gateway com dados de Autenticação

5.3. Resultados

Todas as operações foram realizadas através da API *devices* da *Cloud* com auxílio do Postman², ferramenta para desenvolvimento e testes de APIs.

²www.getpostman.com

Em primeiro lugar, realizamos operações para ligar o ventilador. Ambas foram requisições do tipo PUT para a API da *Cloud*, passando como cabeçalho da requisição as informações de autenticação obtidas no processo de pareamento e mostradas na Figura 7. No corpo da mensagem, passamos um JSON com a operação `set_data` para o sensor com id 1 (ventilador). Em seguida é possível verificar status do sensor 1 através do log do *Gateway*, conforme Figura 8.



(a) Operação: Ligar Ventilador

```
Oct 15 22:46:34 knot daemon.info knotd[238]: WS JSON TX: [ "update", { "get_data": [ ], "uuid": "64ed82d8-5200-431b-a3f2-ece937290000" } ]
Oct 15 22:46:34 knot daemon.info knotd[238]: WS TX83 bytes
Oct 15 22:46:34 knot daemon.info knotd[238]: KNOT OP: 0x20 LEN: 02
Oct 15 22:46:34 knot daemon.info knotd[238]: sensor:1, unit:0, value_type:3
Oct 15 22:46:34 knot daemon.info knotd[238]: JSON: { "sensor_id": 1, "value": true }
Oct 15 22:46:34 knot daemon.info knotd[238]: WS TX149 bytes
Oct 15 22:46:34 knot daemon.info knotd[238]: WS JSON_RX 4 = [{"config":{"uuid":"64ed82d8-5200-431b-a3f2-ece937290000","online":true,"type":"KNOTDevice",
"name":"KNOTThing","owner":"71061601-4729-4348-8130-216a24c90000","socketId":"DHwbZfu0zhfwnmGqAACG","lpAddress":"","ff
Oct 15 22:46:35 knot daemon.info knotd[238]: WS JSON_RX 4 = [ ]
Oct 15 22:46:35 knot daemon.info knotd[238]: WS JSON_TX [ "device", { "uuid": "64ed82d8-5200-431b-a3f2-ece937290000" } ]
Oct 15 22:46:35 knot daemon.info knotd[238]: WS TX67 bytes
Oct 15 22:46:35 knot daemon.info knotd[238]: WS JSON_RX 4 = [{"fromUuid":"64ed82d8-5200-431b-a3f2-ece937290000","from":{"uuid":"64ed82d8-5200-431b-a3f2-ece937290000","online":true,"type":"KNOTDevice","name":"KNOTThing","owner":"71061601-4729-4348-8130-216a24c90000","s
```

(b) Atualização do Ventilador

Figura 8. Operações através da *Cloud*

Por fim, desenvolvemos uma aplicação utilizando a linguagem de programação NodeJS para ligar ou desligar o ventilador de acordo com os dados obtidos no sensor de presença. A aplicação realiza as operações através de uma conexão *websocket* com a Meshblu. A princípio ele faz uma subscrição para receber todas as mensagens correspondentes ao dispositivo pareado, de acordo com sua UUID. Caso uma mensagem chegue indicando que o status do sensor_2 (sensor de presença) seja *true*, a aplicação chama um evento *update* para ligar o sensor_1 (ventilador). Isso é mostrado na Figura 9.

```
Ventilador Desligado!
{ sensor_id: '1',
  value: 'false',
  uuid: '64ed82d8-5200-431b-a3f2-ece937290000',
  timestamp: '2017-10-15T23:59:02.165Z',
  _id: '59e3f64648fef97b679a33e2' }
Ventilador Desligado!
{ sensor_id: '1',
  value: 'false',
  uuid: '64ed82d8-5200-431b-a3f2-ece937290000',
  timestamp: '2017-10-15T23:59:02.165Z',
  _id: '59e3f64648fef97b679a33e2' }
Ventilador Desligado!
{ sensor_id: '2',
  value: 'true',
  uuid: '64ed82d8-5200-431b-a3f2-ece937290000',
  timestamp: '2017-10-15T23:59:02.486Z',
  _id: '59e3f64648fef97b679a33e3' }
Ventilador Ligado!
```

(a) Aplicação detecta presença

```
{ sensor_id: '2',
  value: 'false',
  uuid: '64ed82d8-5200-431b-a3f2-ece937290000',
  timestamp: '2017-10-15T23:58:56.630Z',
  _id: '59e3f64048fef97b679a33db' }
Ventilador Desligado!
{ sensor_id: '1',
  value: 'false',
  uuid: '64ed82d8-5200-431b-a3f2-ece937290000',
  timestamp: '2017-10-15T23:58:57.812Z',
  _id: '59e3f64148fef97b679a33dc' }
Ventilador Desligado!
```

(b) Nenhuma pessoa presente no ambiente

Figura 9. Aplicação: presença e ventilador

6. Conclusão e Trabalhos Futuros

Como podemos verificar através dos resultados, a metaplataforma KNoT pode ser utilizada para viabilizar aplicações de Casas Inteligentes com uma única plataforma de *Cloud* integrada. Os resultados foram obtidos através de operações realizadas manualmente nas APIs disponibilizadas pela plataforma Meshblu, e também através de uma mini-aplicação implementada em NodeJS, demonstrando a capacidade para o desenvolvimento de aplicações robustas que possam atender vários cenários.

Os três componentes da arquitetura do KNoT funcionaram bem em conjunto. O dispositivo realizou o controle e/ou sensoriamento do ventilador e sensor de presença, enviando dados em um intervalo de tempo e respondendo as solicitações feitas através da *Cloud*. Apesar das funcionalidades fundamentais de cada componente funcionarem, algumas dificuldades foram encontradas:

- Integração entre interface web (*Gateway*) e serviços de comunicação com o dispositivo apresentando falhas. Toda configuração foi feita manualmente através dos arquivos de configuração.
- Ao longo da realização das demonstrações foi constatado que o serviço de Fog (instância do Meshblu no *Gateway*) não estava funcionando. O que evitou a realização de operações *on-premise* (localmente).

Para trabalhos futuros, pretendemos desenvolver aplicações com mais de uma plataforma de *Cloud* integrada e em múltiplos domínios. Além disso, também pretendemos implementar uma aplicação totalmente fim a fim, com o desenvolvimento de uma aplicação Android para operar em cima das plataformas, através da biblioteca de abstração (descrita na arquitetura).

Referências

- [Atzori et al. 2015] Atzori, L., Iera, A., and Morabito, G. (2015). The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259.
- [Gartner 2017] Gartner (2017). Gartner's 2016 hype cycle for emerging technologies. <http://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>. Accessed: 2017-10-8.
- [Manyika 2015] Manyika, J. (2015). *The Internet of Things: Mapping the value beyond the hype*.
- [Mineraud et al. 2016] Mineraud, J., Mazhelis, O., Su, X., and Tarkoma, S. (2016). A gap analysis of internet-of-things platforms. *Computer Communications*, 89:5–16.
- [Runeson and Höst 2009] Runeson, P. and Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164.
- [Weiser 1991] Weiser, M. (1991). The computer for the 21 st century. *Scientific american*, 265(3):94–105.
- [Weiser et al. 1999] Weiser, M., Gold, R., and Brown, J. S. (1999). The origins of ubiquitous computing research at parc in the late 1980s. *IBM systems journal*, 38(4):693–696.

PROJETO DE EXTENSÃO DE MATEMÁTICA: AS DIFICULDADES DA APRENDIZAGEM EM MATEMÁTICA

Cleber Gabriel Torres¹, Douglas Albuquerque¹, Jailson Ferreira de Carvalho¹

¹Universidade de Pernambuco – Campus Governador Miguel Arraes de Alencar (FACITEC) CEP 55002-971, Rodovia BR 104, Km 62 – Nova Caruaru, Caruaru-PE

cleber.gabriel@outlook.com.br, douglas-albuquerque@hotmail.com,
jailsonfc.adm@hotmail.com

ABSTRACT *This extended abstract represents an elucidation of importance, and the functioning of the Mathematics Extension Project of the Administration course of the University of Pernambuco on the campus of the city of Caruaru-PE.*

RESUMO este resumo estendido representa a elucidação da importância, e o funcionamento do Projeto de Extensão de Matemática, do curso de Administração da Universidade de Pernambuco do campus da cidade de Caruaru-PE.

INTRODUÇÃO

O Projeto de Extensão de matemática retrata uma ação referente ao mau desempenho de muitos dos estudantes de graduação do curso de Administração da Universidade de Pernambuco sobre a disciplina de matemática, e suas implicações acerca do aprendizado relacionado com as dificuldades e facilidades enfrentadas pelos mesmos. Pois, ao ingressar na iniciação científica, a matemática ganha contornos de aplicação na vida profissional, então, há uma seleção entre aqueles que optam por áreas afins e não com a matéria, tendo a segunda opção como a preferida pela maioria. Logo, de acordo com a lógica da oferta e demanda, há uma escassez por pessoas que se identificam com o setor, e devido a sua aplicação universal, somos impedidos de evoluirmos em diversos segmentos da tecnologia e do conhecimento por falta de profissionais e pesquisadores do assunto.

REFERENCIAL TEÓRICO

“Na universidade, a matemática adquire um caráter distinto, é cobrado dos alunos uma experiência anterior que eles não tem” (BROLEZZI, 2004,p. 1), no ensino superior entende-se que o aluno já deve ter um certo nível em conhecimentos matemáticos,

porém não é o que acontece na prática, principalmente com ex-alunos de escolas da rede pública de ensino.

Segundo Leinold (1988), desde o entendimento da economia, dosagem de medicamentos a expansão da população de bactérias na biologia, encontra-se alguma aplicação de conceitos matemáticos. Logo, de acordo com o pensamento do autor, a matemática está evidenciada em vários setores de nossas vidas, em muitas das vezes passada por despercebida, por ser uma disciplina onipresente, seu aprendizado não deveria apresentar tanta dificuldade como se constata.

Cury (2009) relata que entre 1999 e 2009 as dificuldades de aprendizagem principalmente de cálculos se tornaram mais frequente e preocupante, ficando mais evidente a falta de conhecimentos prévios dos assuntos ou a compreensão equivocada em níveis anteriores. Sabemos que hoje tendo se passado oito anos, infelizmente o cenário continua o mesmo.

CONCLUSÃO

O projeto de extensão de matemática da Universidade de Pernambuco do campus Caruaru, tem o objetivo de nivelar os alunos de acordo com a exigência necessária para o curso, através da revisão de assuntos do ensino médio, permitindo que os mesmos tenham conteúdo necessário para desenvolver as atividades propostas pelas disciplinas que envolvam equações. No curso de Administração não temos apenas uma matéria que envolva cálculos, pelo contrário, ao longo do curso nos deparamos com várias, a exemplo da matemática financeira, contabilidade gerencial, estatística entre outras.

O Projeto de Extensão de Matemática na UPE é dedicado para todos os alunos que almejam aprender mais da matéria, acontecendo uma vez por semana, sendo ministrado pelo doutorando Alex Borges, membro do quadro de funcionários da instituição. Ocorrendo as sextas-feiras, das 17:30 as 18:30, antecedendo as aulas da cadeira de matemática do segundo período. Os alunos ainda contam com o auxílio de três monitores, estudantes do quarto período, que realizam monitorias nas quintas-feiras das 18:30 às 17:30. E ao final do projeto, os alunos que obtiveram um mínimo de 75% de frequência nas aulas, ganhará um certificado de participação com carga horária de 20h, servindo como horas complementares obrigatória do curso.

REFERÊNCIAS

- BROLEZZI, A. C. (2004). **Mudanças na Matemática da Escola Básica para o Ensino Superior: reflexo no uso de História da Matemática.** In: VII EPEM - Encontro Paulista de Educação Matemática, São Paulo. Anais do VII EPEM. Disponível em: <http://miltonborba.org/CD/Interdisciplinaridade/Anais_VII_EPEM/grupos_trabalho/gdt08-Brolezzi2.doc> Ou: <http://www.ime.usp.br/~brolezzi/publicacoes/gdt08-Brolezzi2.doc>.
- CRUZ, Fernando (2013). **Escolas modelo podem ajudar a melhorar ensino de matemática.** Disponível em: <<https://educacao.uol.com.br/noticias/2013/03/06/escolas-modelo-poderao-ser-usadas-para-melhorar-ensino-de-matematica.htm>>. Acesso em: 10 out 2017.
- CURY, H. N (2009). **Pesquisas em análises de erros no ensino superior: retrospectiva e novos resultados.** In: FROTA, M. C. R., NASSER, L. (Org.). Educação matemática no ensino superior: pesquisas e debates. Recife: SBEM. 265p.
- FONSECA, Ricardo F. CASTRO SILVA, Alex W (2012). **A Relação Entre o curso de administração, Os Acadêmicos e a Disciplina de Matemática: Uma reflexão.** Convibra Administração. [S.I]. Disponível em: <http://www.convibra.com.br/upload/paper/2012/31/2012_31_4363.pdf>. Acesso em: 09 de outubro de 2017.
- FUENTES, Verônica L. P. LIMA, Ronaldo. GUERRA, Diego S. **Atitudes em Relação a Matemática em Estudantes de Administração.** Revista Semestral da Associação Brasileira de Psicologia Escolar e Educacional (ABRAPEE). Uberlândia. Volume 13, Número 1, Janeiro/Junho de 2009. 133-141.

O Avanço que a Monitoria Proporciona ao Aprendizado de Cálculo

Matheus Barboza¹, Mirele Moutinho Lima¹

¹Universidade de Pernambuco (UPE)
Caruaru– PE – Brasil

{matheus.barboza, mirele.moutinho}@upe.br

Abstract. *Monitoring is one of the fundamental activities to generate good performance during higher education, especially in the areas of Exact Sciences. Differential and Integral Calculus 1 and 2 are curricular components that require a greater degree of dedication both of the student and of those who teach, so it is necessary the support of the monitor for a better performance and a greater chance of success in the approval of the discipline.*

Resumo. *A monitoria é uma das atividades fundamentais para gerar um bom desempenho durante o ensino superior, principalmente tratando-se de disciplinas da área de Ciências Exatas. Cálculo Diferencial e Integral 1 e 2 são componentes curriculares que exigem um grau maior de dedicação tanto do aluno, quanto de quem leciona, por isso é necessário o apoio do monitor para um melhor desempenho e maior chance de sucesso na aprovação da disciplina.*

1. Introdução

A finalidade deste trabalho é mostrar o porquê de a atividade de monitoria ser fundamental para o desenvolvimento dos discentes que estão iniciando o aprendizado. Segundo Reis (2005, p.1), as causas mais frequentes para a rejeição do aprendizado em fundamentos matemáticos, desde o ensino fundamental, podem ser dadas por vários motivos. A falta de motivação dos professores e alunos, o rigor matemático, a ausência de ligação entre a matemática escolar e o cotidiano dos alunos, experiências negativas envolvendo esta matéria e a maneira como os professores desenvolvem suas atividades.

Consequentemente isso é sentido por anos seguintes, no ensino universitário, e o índice de reprovação em Cálculo Diferencial e Integral 1 e 2 no curso de Sistemas de Informação da Universidade de Pernambuco pode ser alto, justamente pelo déficit em matemática que os discentes carregam durante uma grande parte da vida acadêmica dos primeiros anos de vida.

2. Métodos

Uma solução que vem dando resultados satisfatórios para esse problema são as monitorias nas componentes curriculares dessa área. O projeto **A Inovação Tecnológica**

no Aprendizado de Cálculo Diferencial e Integral promove apoio do discente monitor aos alunos de Cálculo Diferencial e Integral 1 e 2, em aulas com o intuito de revisar os assuntos ministrados em sala, oferecer suporte às resoluções das listas de exercícios e principalmente esclarecer dúvidas. A principal característica é a dinamicidade das aulas, que acompanham a necessidade e o ritmo de aprendizagem da turma. Além disso, a interação entre aluno e monitor também tem o apoio de ferramentas e aplicativos cujo recurso base é a internet, como e-mail, Facebook e WhatsApp, algo que foi parte fundamental para bons resultados, principalmente quando se tratando de esclarecer dúvidas das questões das listas de exercício.

No começo do ano letivo, na semana de acolhimento, foi realizado o curso de Pré-Cálculo, o qual serviu para preparar e dar suporte aos alunos antes mesmo do início propriamente dito dos conteúdos programáticos, amenizando os efeitos negativos da dificuldade em matemática. Foram abordados assuntos básicos do ensino médio como, funções de primeiro e segundo grau, funções trigonométricas, expressões algébricas, produtos notáveis e fatoração, assuntos que são importantes para ter um bom nível de compreensão da matéria. As aulas de monitoria acontecem uma vez na semana, com a possibilidade de mais um encontro nas semanas que antecedem as avaliações, ou sempre que necessário, e visam o aprendizado de cada tema abordado pela professora.

3. Resultados

Ao fazer-se uso da metodologia de estudo presente na componente de Cálculo 1 no semestre inicial do ano de 2017, foram encontrados resultados considerados interessantes e de forte teor de impacto, em defesa da existência do projeto, devido ao fato de que, conforme explícito no gráfico 1, quase 80% dos alunos que frequentaram a monitoria obtiveram êxito em suas consequentes aprovações, ao passo que, estudantes que não frequentavam as monitorias tiveram uma taxa altíssima de desistência da disciplina, além de apresentar valores elevados de reprovações, conforme elucidado no gráfico 2.

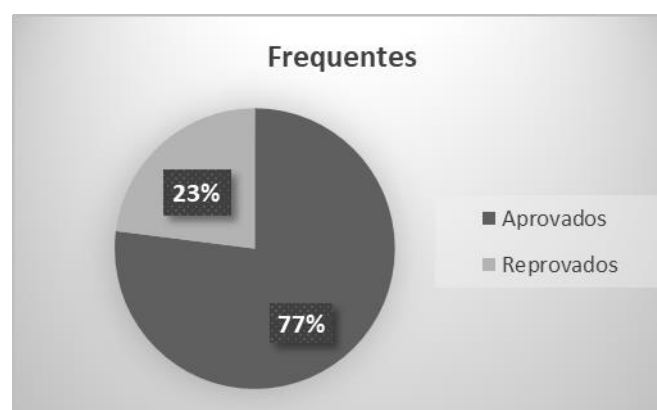


Figura 1: Dados estatísticos dos discentes que frequentaram a monitoria na disciplina de Cálculo 1

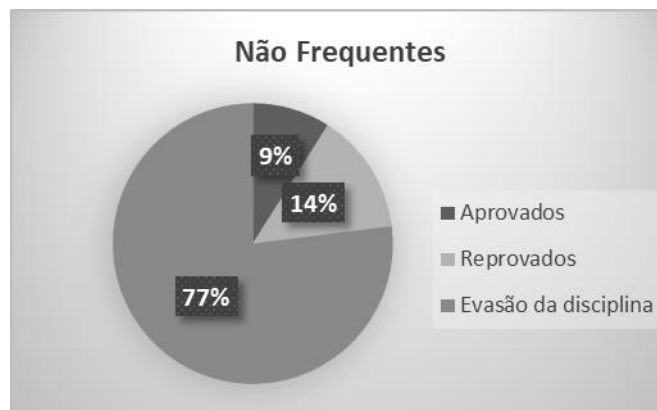


Figura 2: Dados estatísticos dos discentes que não frequentaram a monitoria na disciplina de Cálculo 1.

4. Conclusões

A partir desses resultados, pode-se concluir que o percentual de aprovação dos alunos que compareceram às aulas de monitoria de Cálculo 1 é muito maior se comparado aos que não frequentam, por consequência disso, os mesmos alunos que frequentarão as aulas de Cálculo 2, provavelmente terão uma melhor base e maior chance de aprovação nessa próxima etapa. Assim, fica claro que o projeto **A Inovação Tecnológica no Aprendizado de Cálculo Diferencial e Integral** foi bastante eficiente em atender sua proposta e a necessidade de continuidade dos projetos de monitoria é imprescindível para o bom rendimento dos alunos.

Referências

- Reis, Leonardo Rodrigues. (2005). “Rejeição à matemática: causas e formas de intervenção”. Disponível em: <https://www.ucb.br/sites/100/103/TCC/12005/LeonardoRodriguesdosReis.pdf>
- Souza, Paulo Rogerio Areias de. (2017). “A importância da monitoria na formação de futuros professores universitários”. In: Âmbito Jurídico. Disponível em: http://www.ambito-juridico.com.br/site/index.php?n_link=revista_artigos_leitura&artigo_id=5990
- Souza, R. O; Gomes, A. R. (2015) “A eficácia da Monitoria no Processo de Aprendizagem visando a Permanência do Aluno na IES”. In: Revista Interdisciplinar. do Pensamento Científico, Nº 2, Vol. 1, Art. nº 16, Julho/Dezembro. Disponível em: <http://reinpec.srvroot.com:8686/reinpec/index.php/reinpec/article/download/106/42>

Bancos de Dados: Conectando Tradição e Atualidade no Desenvolvimento do Currículo do Discente de Sistemas de Informação

Sóstenes Gomes de Souza¹, Elyda Laisa Soares Xavier Freitas^{1,2}

¹Sistemas de Informação - Universidade de Pernambuco (UPE)
Caruaru- PE – Brasil

²Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Recife – PE - Brasil

sostenesg7@gmail.com, elyda.freitas@upe.br

Abstract. *Monitoring is a service of pedagogical support that contributes to the good performance of students. It is essential, therefore, that Higher Education institutions provide this extra support to them. This article presents the execution's experience of the project of monitoring to the Databases' discipline, supported by the University of Pernambuco.*

Resumo. *A monitoria é um serviço de apoio pedagógico que contribui para o bom aproveitamento dos discentes. É importante, assim, que as instituições de ensino superior forneçam esse suporte extra aos mesmos. Este artigo apresenta a experiência de execução do projeto de monitoria para a disciplina de Banco de Dados, apoiado pela Universidade de Pernambuco.*

1. Introdução

“A monitoria docente é um serviço de apoio pedagógico oferecido aos alunos interessados em aprofundar conteúdos, bem como solucionar dificuldades em relação à matéria trabalhada em aula” (HAAG et al., 2008). Além do benefício ao aluno monitorado, Costa e Baltar (2015) destacam que a monitoria proporciona ao monitor o aprimoramento de sua prática docente, por meio do auxílio e acompanhamento constante. Nesse contexto, este artigo apresenta a experiência de monitoria da disciplina de Banco de Dados 2 em um curso de Sistemas de Informação, objetivando fornecer informações acerca das atividades realizadas, incluindo a ministração das aulas, e resultados do desempenho escolar dos estudantes, bem como possíveis pontos de melhoria.

2. Andamento das Aulas

O projeto teve seu desenvolvimento na Universidade de Pernambuco – Campus Caruaru, durante o ano letivo de 2017. As monitorias tiveram uma periodicidade de um encontro presencial semanal, com duração média de quatro horas. Pode-se dizer que o intuito principal das aulas era envolver os monitorados com problemáticas do mundo real, fazendo uso de linguagem mais direta e clara, da realização de questionamentos e do uso intensivo de exercícios, como forma de fixação de conteúdo e identificação de falhas no aprendizado. Fora das aulas no âmbito universitário, houve a necessidade de utilização de ambientes virtuais colaborativos como *Facebook* e *Whatsapp*, *Google Hangouts* e *Google Docs*, para acompanhamento, esclarecimento de dúvidas, compartilhamento de material auxiliar e manutenção do relacionamento com os estudantes. Mesmo com a

possibilidade de criação de salas de aulas virtuais, em ambientes próprios para a realização dessas atividades, como o *Google Classroom*, as ferramentas utilizadas foram escolhidas por estarem em contato direto e constante com os alunos, sem a necessidade de acesso a outros ambientes.

3. Resultados e Discussão

Ao término das atividades de monitoria, referentes ao primeiro semestre letivo, o efeito do trabalho realizado pôde ser analisado por meio de questionários e pela avaliação dos resultados dos estudantes. Esses resultados são apresentados nesta seção.

3.1. Quanto à Opinião dos Alunos

Para entender melhor o papel e a importância da monitoria no aprendizado dos discentes, ao término do semestre letivo um questionário foi elaborado e disponibilizado para que os estudantes realizassem a avaliação de forma anônima. Os resultados obtidos, dos oito alunos que responderam, estão apresentados abaixo, na Figura 1, com base nos seguintes questionamentos:

- Qual o nível de dificuldade da disciplina Bancos de Dados 2?
- Em que escala você avalia a importância da monitoria (inclusive documentos e informações fornecidas, numa escala de 0 a 10)?
- Como você avalia a importância da utilização de ferramentas colaborativas (*Facebook*, *Whatsapp*, *Google Hangouts* e *Google Docs*), para a ministração da monitoria e tira-dúvidas?

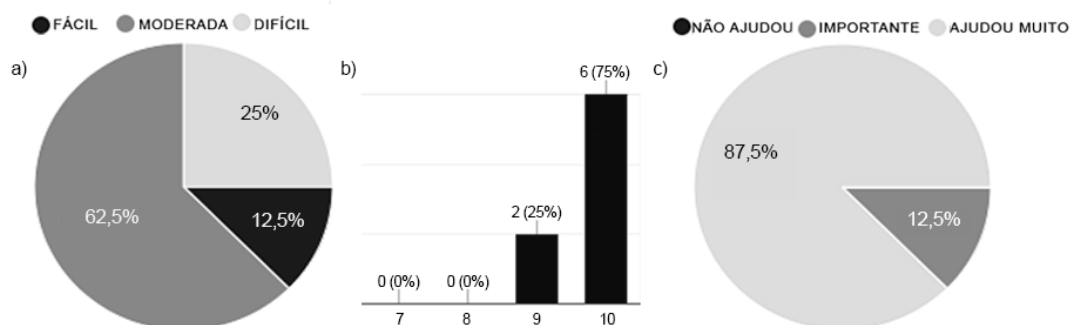


Figura 1. a) Avaliação do nível de dificuldade da disciplina Bancos de Dados II; b) Avaliação da importância da monitoria e materiais fornecidos; c) Avaliação da importância da utilização de ferramentas colaborativas.

Os resultados expõem a opinião dos alunos que, em sua maioria, consideram Banco de Dados 2 uma disciplina de moderada (62,5%) a difícil (25%). Ademais, a monitoria foi avaliada positivamente, uma vez que todos os alunos respondentes a avaliaram com 9 ou 10 pontos numa escala de importância de 0 a 10. Por fim, o uso de ferramentas colaborativas também foi avaliado positivamente, com 87,5% dos alunos concluindo que as mesmas os ajudaram durante a monitoria.

3.2. Quanto ao Aproveitamento Escolar dos Alunos

No que se refere ao 1º EE (Exercício Escolar), a Figura 3a mostra que dos 25 (vinte e cinco) alunos cursando a disciplina, 16% deles alcançaram nota igual ou acima da média. Além disso, 20% dos alunos faltou no dia da aplicação do 1º EE. Os demais, 64%,

obtiveram média menor que 7 (sete). Sobre o 2º EE, observa-se que o desempenho dos alunos melhorou, como apresentado na Figura 3b: 40% alcançou nota igual ou acima da média. Ainda, 40% dos alunos faltou no dia da aplicação do 2º EE e apenas 20% dos alunos não atingiu a média.

Por fim, a Figura 3c mostra ainda que dos 11 alunos que prestaram o exame final, apenas 1 (um), ou seja, 9%, não conseguiu obter a nota necessária para passar na disciplina. Sendo assim, dos 25 alunos que cursaram a disciplina, excetuando os reprovados por falta, apenas aluno 1 (um) não obteve média para ser aprovado.

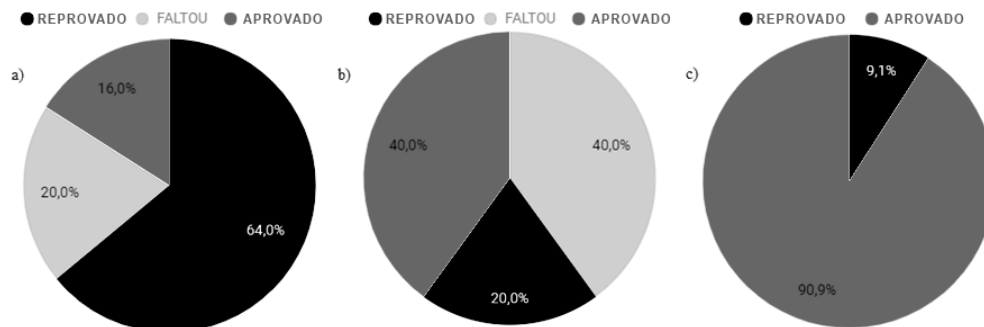


Figura 2. Índice de aprovação dos alunos a) no 1º Exercício Escolar e b) no 2º Exercício Escolar e c) no exame final.

4. Conclusão

Com base nos resultados obtidos e aqui expostos, pode-se concluir que o trabalho ora realizado teve contribuição para o desenvolvimento dos alunos durante a disciplina de Banco de Dados 2, ratificando a monitoria como elemento importante no auxílio à obtenção do conhecimento. Apesar de um baixo índice de aprovação na primeira avaliação, o desempenho dos alunos melhorou ao longo do semestre de modo que os alunos que optaram por seguir na disciplina obtiveram em sua grande maioria resultados satisfatórios. Além disso, pela aplicação do questionário, verificou-se que a avaliação da monitoria, por parte dos monitorados, também foi positiva. Também, o resultado do projeto de monitoria é considerado construtivo dado que no semestre anterior, quando não havia monitor para auxílio aos estudantes, a disciplina de Bancos de Dados 2 obteve um índice de reprovação de 23% (excetuando os reprovados por falta), contra 9% no semestre corrente, quando houve atuação do monitor.

O alto índice de evasão é um fator preocupante, cabendo uma atenção mais direcionada nos próximos semestres, a fim de encontrar possíveis fatores que o justifiquem. Por fim, o presente projeto segue em desenvolvimento, sendo aplicado com uma nova turma, durante o segundo semestre do ano corrente (2017.2), na disciplina de Bancos de Dados 1.

5. Referências

- COSTA, J. S.; BALTAR, S. L. S. M. A. A importância e concepção da monitoria de estágio supervisionado para alunos do curso de licenciatura em Biologia. *Revista Iniciação e Formação Docente*. v. 1, n. 2, 2015.
- HAAG, G. S.; KOLLING, V.; SILVA, E.; MELO, S. C.B.; PINHEIRO, M. Contribuições da Monitoria no processo de ensino-aprendizagem em enfermagem. *Revista Brasileira de Enfermagem*. v.61, n. 2, p. 215-220. Março-abril de 2008.

Integração de infraestrutura de computação em nuvem e IoT para fornecimento de aplicações

Matheus Felipe Ferreira da Silva Lisboa Tigre¹, Patricia Takako Endo¹

¹Universidade de Pernambuco (UPE)
Caruaru – PE – Brasil

matheus0906.mhci@gmail.com, patricia.endo@upe.br

Resumo. *Internet das coisas e computação em nuvem são duas áreas que cada vez mais vem sendo trabalhadas de forma conjunta, visando utilizar os recursos de ambas para o bem comum. Dentre a inúmera quantidade de vantagens que essa união oferece, está a captura, processamento e distribuição de dados obtidos de sensores através da infraestrutura da nuvem para aplicações utilizadas por interessados nessas informações. Isso dá origem ao que pode ser chamado de Sensing-as-a-Service. Nesse contexto, existem várias aplicações possíveis com a utilização dessa estrutura e que podem beneficiar diversos tipos de stakeholders. Unindo computação em nuvem e IoT pode trazer benefícios tanto de um para o outro como para a disponibilização de dados pela rede.*

1. Introdução

Nuvem e IoT (*Internet of Things*), apesar de englobarem conceitos diferentes, apresentam características que podem auxiliar um ao outro em contextos específicos. Os dispositivos utilizados em IoT possuem pouca capacidade de armazenamento e processamento, o que acaba gerando problemas como performance, segurança e etc. Já computação em nuvem provê acesso à grande quantidade de armazenamento e processamento poderoso de maneira escalável e por demanda. Logo, uma integração entre os dois pode ajudar a solucionar grande parte desses problemas [Babu et al. 2015]. Se por um lado, IoT utiliza os recursos ilimitados de maneira virtual, como armazenamento, processamento e comunicação providos pela nuvem, esta pode estender sua área de abrangência, provendo novos serviços utilizando objetos da vida real que são as "coisas" da IoT. Essa integração pode ser chamada de *CloudIoT* [Botta et al. 2016].

Este artigo tem como objetivo apresentar um dos serviços possibilitados pela *CloudIoT* que é conhecido como *Sensing as a service* (SenaaS) e descrever algumas aplicações que o utilizam.

2. Definições sobre Sensing-as-a-Service

Um exemplo de utilização da *CloudIoT* pode ser a *Sensor Cloud*, que utiliza sensores como uma interface entre o mundo real e o mundo virtual, a infraestrutura de computação em nuvem como um recurso de processamento de dados e a Internet como meio de comunicação. A *Sensor Cloud* reúne uma grande rede de sensores com aplicações de *sensing* onde é possível processar, analisar e armazenar os dados obtidos pelos sensores [Rao et al. 2012].

Em [Abdelwahab et al. 2014], a utilização de uma rede de sensores aliada à computação em nuvem é chamada de "*cloud assisted remote sensing*" (CARS). Ela permite que os dados sejam obtidos de vários lugares diferentes, compartilhados para diversos usuários, acessados de qualquer parte de mundo, utilizar recursos por demanda (de acordo com sua preferência) e pagar somente pelo que utilizou. Quando isso é oferecido como um serviço, dá origem ao que se chama de SenaaS [Abdelwahab et al. 2014]. Trata-se de mais um serviço presente na nuvem, onde os sensores são vistos como coletores de dados e seus proprietários disponibilizam os dados coletados para os usuários ou para aplicações que necessitem das informações detectadas pelos dispositivos [Mathew et al. 2016].

Uma das vantagens da utilização de SenaaS é que os custos com a manutenção da infraestrutura de sensores, como consertos e bateria, podem vir a serem divididos entre os proprietários e os consumidores dos dados, tendo em vista que, esses últimos irão pagar pela utilização dos sensores de acordo com o que é utilizado.

3. Aplicações

Uma área que pode ser auxiliada pela *CloudIoT* é a de *healthcare*, mais especificamente, quando se trata de *health monitoring* (HM). Em [Neagu et al. 2017] é destacado como o monitoramento da saúde de idosos pode se beneficiar da *CloudIoT*, ajudando no tratamento de idosos em suas residências como alerta da hora de medicamentos e disponibilização em tempo real do estado de saúde do paciente. Em [Islam et al. 2015], são exemplificados algumas aplicações de IoT para HM, como: monitoramento de glicose (dispositivos de captura e medição da glicose no sangue), monitoramento de batimentos cardíacos (eletrodos conectados em uma rede sem fio), monitoramento de pressão sanguínea (equipamento de medição de pressão conectado na rede que medindo e transmitindo as informações) e monitoramento de temperatura corporal (sensor capturando a temperatura do corpo do paciente).

Uma outra situação que pode ser facilitada e melhorada com a utilização da *CloudIoT* é a coleta e o transporte do lixo. Com o crescimento das cidades, isso acabou virando um problema. Uma solução para esse problema pode ser observada em [Aazam et al. 2016]. Nele os autores apresentam um monitoramento inteligente de lixo, onde as lixeiras são equipadas com sensores que medem sua capacidade. Mas, como é dito no trabalho, para se ter um controle de lixo inteligente de fato, é necessário muito mais do que disponibilizar dados da capacidade da lixeira. Por isso, o objetivo do trabalho é utilizar a nuvem para incrementar esse cenário, disponibilizando informações para os interessados no controle de lixo (como companhias de reciclagem, indústria da saúde, da coleta de lixo, entre outros). As vantagens que podem ser adquiridas com um controle inteligente de lixo vão desde um melhor horário de coleta, otimização de da rota no transporte, reciclagem e descarte, entre outras coisas.

4. Conclusão

Nos dias de hoje, o controle automatizado por meio de sensores conectados à rede é cada vez mais recorrente. Todos os dias, uma enorme quantidade de dados é adquirida de diferentes tipos de dispositivos ao redor do mundo, dispositivos estes que não possuem capacidade de armazenamento e processamento necessários para lidar com tamanha quan-

tidade de informação. É neste contexto onde entra a nuvem, resolvendo esse problema oferecendo processamento e armazenamento ilimitados.

Mas, dos mesmo modo que existe muita informação sendo adquirida, existem muitos interessados em obter acesso a ela, logo, é de extrema importância que seja encontrada uma maneira de disponibiliza-la para todos os *stakeholders*. Isso é uma tarefa para a computação em nuvem disponibilizando o compartilhamento dos dados como um serviço através de sua infraestrutura, beneficiando, assim, todos os envolvidos, tanto os interessados, como os proprietários dos sensores que irão conseguir dividir os custos com a rede de sensores que possuem com os usuários que adquirem os serviços.

Outra vantagem é a comunicação em tempo real. Os usuários podem monitorar aquilo que desejarem da maneira mais rápida e eficiente, obtendo informação em tempo real que pode os ajudar de acordo com suas preferências. Ou seja, computação em nuvem e IoT são duas tecnologias que juntas podem evoluir bastante a disponibilização de informação pela rede.

Referências

- Aazam, M., St-Hilaire, M., Lung, C.-H., e Lambadaris, I. (2016). Cloud-based smart waste management for smart cities. In *Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2016 IEEE 21st International Workshop on*, pages 188–193. IEEE.
- Abdelwahab, S., Hamdaoui, B., Guizani, M., e Rayes, A. (2014). Enabling smart cloud services through remote sensing: An internet of everything enabler. *IEEE Internet of Things Journal*, 1(3):276–288.
- Babu, S. M., Lakshmi, A. J., e Rao, B. T. (2015). A study on cloud based internet of things: Clouddiot. In *Communication Technologies (GCCT), 2015 Global Conference on*, pages 60–65. IEEE.
- Botta, A., De Donato, W., Persico, V., e Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future Generation Computer Systems*, 56:684–700.
- Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., e Kwak, K.-S. (2015). The internet of things for health care: a comprehensive survey. *IEEE Access*, 3:678–708.
- Mathew, S. S., Atif, Y., e El-Barachi, M. (2016). From the internet of things to the web of things—enabling by sensing as-a service. In *Innovations in Information Technology (IIT), 2016 12th International Conference on*, pages 1–6. IEEE.
- Neagu, G., Preda, Ș., Stanciu, A., e Florian, V. (2017). A cloud-iot based sensing service for health monitoring. In *E-Health and Bioengineering Conference (EHB), 2017*, pages 53–56. IEEE.
- Rao, B. P., Saluia, P., Sharma, N., Mittal, A., e Sharma, S. V. (2012). Cloud computing for internet of things & sensing based applications. In *Sensing Technology (ICST), 2012 Sixth International Conference on*, pages 374–380. IEEE.

Resultados Preliminares um Mapeamento Sistemático: EaD com Suporte em PBL no Contexto da Computação

Silvano R. de A. Neto¹, Wylliams Barbosa Santos¹, Aline S. Gomes¹

¹GRupo de Estudos Avançados em Tecnologia da Informação e Comunicação (GREAT)
Universidade de Pernambuco (UPE), Caruaru – PE – Brazil

{silvano.neto, wbs}@upe.br, aline.dasilvagomes12@gmail.com

Resumo. Contexto: A necessidade de trabalhar resolução de problemas reais em aula através de interações sem limites temporais e geográficos tornam a integração entre a metodologia PBL e a modalidade EaD uma tendência. **Objetivo:** Este trabalho tem como objetivo identificar, quantificar e analisar o conhecimento produzido sobre o tema no contexto da computação. **Método:** Um Mapeamento Sistemático da Literatura está sendo conduzido, a fim de revelar o conhecimento produzido através dos estudos primários. **Resultados:** Buscas experimentais e avaliações primárias dos resultados de busca demonstram números importantes para o andamento do mapeamento. **Conclusão:** Direcionamentos estão sendo explorados para trabalhos futuros.

1. Introdução

A literatura aborda uma série de estratégias de ensino e exemplos de implantação destas, cada uma no intuito de atender as necessidades metodológicas daquela determinada área ou disciplina. Um dos desafios é englobar o uso do conhecimento em sala de aula na resolução prática de problemas do cotidiano, tanto para a integração do aluno no mercado de trabalho quanto para o desenvolvimento de pontes com a pesquisa.

Além da Aprendizagem Baseada em Problemas (PBL), os avanços da tecnologia da informação também tem colaborado com a realidade da educação, flexibilizando o processo de ensino-aprendizagem que até então exigia tempo e local fixos, através da rede mundial (Internet) e de aplicações voltadas para Ensino a Distância (EaD).

No propósito de conhecer o que há de mais relevante na temática EaD-PBL no contexto do ensino da computação, estão sendo realizados os processos de identificação, quantificação e análise de trabalhos disponibilizados pela academia através de um Mapeamento Sistemático da Literatura (MSL). Para este artigo, serão demonstradas algumas das etapas já realizadas do mapeamento, resultados parciais e expectativas para o seu andamento futuro.

2. Aprendizagem Baseada em Problemas e Ensino a Distância

O ensino voltado para aplicação do conhecimento gera estímulo do pensamento e desenvolve *insight* crítico para resolução de problemas do mundo real [Santos e Soares 2013]. Neste sentido, a PBL é uma metodologia inspirada no construtivismo. Seu objetivo é estimular os estudantes a desenvolver o pensamento crítico, habilidades para resolução de problemas, aprendizagem autônoma, capacidade de comunicação e colaboração [Souza et al. 2014].

Observa-se no entanto, que nem sempre a permanência de alunos e educadores num mesmo local ao mesmo tempo é possível. A necessidade de rompimento das barreiras geográficas no ambiente educacional torna o EaD uma alternativa modal capaz de atender aos anseios de universalização do ensino, como um meio apropriado à contínua atualização dos conhecimentos gerados em grupo [Nunes 1993].

3. Mapeamento Sistemático da Literatura

Muitos campos de pesquisa possuem metodologias de estudos específicas, extensivamente usadas [Petersen et al. 2007]. Com a maturidade de uma área de pesquisa, geralmente há um aumento no número de relatos e resultados. Logo, é importante sumarizar e prover uma revisão. [Petersen et al. 2007] coloca que os principais objetivos do MSL é prover uma completa visão da área de pesquisa: identificar qualidade, tipo de pesquisa e resultados disponíveis na área. Ele também auxilia no mapeamento de frequências das publicações para identificar tendências ao longo do tempo. No intuito de direcionar as etapas de planejamento e métodos aplicados na pesquisa sistemática, um protocolo foi então desenvolvido.

Questões de Pesquisa: Perguntas são formuladas para o direcionamento do mapeamento. O trabalho parte da seguinte questão central: “Qual o estado da arte da aplicação EaD com suporte no método de ensino PBL em disciplinas voltadas para Ciência da Computação?”. Quatro perguntas de pesquisa específicas foram desenvolvidas: (QP1) Quais exemplos da implantação da EaD com suporte em PBL no contexto da Computação temos atualmente na literatura? (QP2) Quais estratégias de integração da PBL nos ambientes EaD foram encontradas na literatura? (QP3) Quais pontos positivos foram detectados mediante a implantação? (QP4) Quais pontos negativos foram detectados mediante a implantação?

Estratégia de Pesquisa: Para a realização das buscas, foram anexadas ao protocolo as seguintes bases automáticas: ScienceDirect¹, IEEE Xplore Digital Library² e ACM Digital Library³. Uma *string* de busca foi construída, de modo a englobar documentos que potencialmente apresentassem respostas significativas às questões de pesquisa: (“*distance education*” OR “*distance learning*” OR “*online learning*” OR “*online education*” OR *e-learning*) AND (“*problem based learning*” OR “*problem based learning*” OR *pbl*) AND (*computer* OR “*engineering computer*” OR “*science computer*” OR “*information technology*” OR “*information system*”).

CrITÉrios de Inclusão: (CI1) Estudos que apresentem relação com a questão (e subquestões) de pesquisa; (CI2) Periódicos e artigos publicados em revistas, conferências ou simpósios; (CI3) Estudos que retratam a experiência de PBL e EaD dentro do contexto de ensino em disciplinas ligadas a Computação; (CI4) Pesquisas relacionadas a aplicação PBL-EaD na educação, nas mais diversas subáreas da Computação.

CrITÉrios de Exclusão: (CE1) Estudos fora do contexto de Ciência da Computação; (CE2) Estudos que não apresentam aplicação do método PBL e do modelo EaD em conjunto; (CE3) Documentos incompletos, rascunhos, documentos de compilação dos anais de conferências (*proceeding*), documentos apenas acessíveis por

¹<http://www.sciencedirect.com>

²<https://ieeexplore.ieee.org/Xplore/home.jsp>

³<http://dl.acm.org>

meio da compra e apresentações em slides; (CE4) Não acessíveis pela Internet; (CE5) Documentos escritos em outras línguas com exceção do inglês; (CE6) Documentos duplicados; (CE7) Documentos que não foram publicados entre os anos de 1997 e 2017.

4. Resultados Parciais e Discussões

Entre os meses de Abril e Maio de 2017, foram realizadas buscas experimentais para a realização de ajustes no protocolo. A partir de Junho e Julho de 2017, os resultados retornados pelos motores de buscas passaram pela primeira etapa de avaliação (análise pelo título e resumo), onde 52 artigos foram selecionados para a segunda etapa (análise pela introdução). Os números estão explicitados na Tabela 1.

Tabela 1. Resultados das buscas e artigos aprovados por motor de busca

Avaliação	ScienceDirect	IEEE Xplore	ACM DL
Documentos retornados	641	233	634
1. Título e Resumo	23	19	10

5. Conclusão e Trabalhos Futuros

O mapeamento está em processo de avaliação dos artigos. A partir da primeira etapa de triagem, cerca de 96,5% dos artigos foram descartados, uma taxa esperada para a especificidade da temática. Durante as buscas experimentais, fase onde todo o protocolo é reajustado com o fim de diminuir o nível de imprecisão dos resultados retornados, previu-se que a taxa de descarte seria ainda maior.

Após a avaliação dos artigos mediante os critérios, dados serão extraídos com a finalidade de responder as questões de pesquisa previamente definidas no protocolo. Essas informações serão de grande importância para o desenvolvimento e aplicação de novos meios de se estudar computação, na expectativa de agregar qualidade e flexibilidade no processo de aprendizagem.

Referências

- Nunes, I. B. (1993). Noções de educação a distância. *Revista educação à distância*, 4(5):7–25.
- Petersen, K., Feldt, R., Mujtaba, S., e Mattsson, M. (2007). Systematic Mapping Studies in Software Engineering. *12th International Conference on Evaluation and Assessment in Software Engineering*, 17(1):1–10.
- Santos, S. C. d. e Soares, F. S. (2013). Authentic assessment in software engineering education based on pbl principles: a case study in the telecom market. In *Proceedings of the 2013 International Conference on Software Engineering*, pages 1055–1062. IEEE Press.
- Souza, A. S., Duran, A., e Vieira, V. (2014). Uma ontologia de domínio para a metodologia de aprendizagem baseada em problemas. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, page 1253.

Problematização através do PBL como suporte ao ensino-aprendizagem: Resultados Parciais de um Mapeamento Sistemático

Aline S. Gomes¹, Wylliams Barbosa Santos¹, Silvano R. de A. Neto¹

¹GRupo de Estudos Avançados em Tecnologia da Informação e Comunicação (GREAT)
Universidade de Pernambuco (UPE), Caruaru – PE – Brazil

{aline.dasilvagomes12@gmail.com, wbs@upe.br, silvano.neto@upe.br}

Resumo. *Contexto:* A necessidade de propostas de metodologias de ensino que proporcionem um maior grau de satisfação e conhecimento por parte dos estudantes. *Objetivo:* Compreender PBL e Metodologia da Problematização como estratégias de ensino-aprendizagem significativas e eficazes, que poderiam ser utilizadas nos diversos níveis de ensino e em múltiplas áreas de conhecimento. *Método:* Por meio de um Mapeamento Sistemático da Literatura, buscou-se refletir acerca da sua importância como método inovador de aprendizagem. *Resultados:* A partir das pesquisas iniciais, foi observada a importância desses métodos como facilitadores em situações de ensino, apresentando resultados propositivos na literatura para a apreensão de aprendizagens significativas, observados por pesquisadores de diversas áreas. *Conclusão:* Espera-se possibilitar a reflexão sobre novas estratégias de aprendizagem com o uso de problematizações e novos direcionamentos para investigações futuras.

1. Introdução

Nos últimos anos, o processo tradicional de ensino em sala de aula tem gerado uma extensa discussão e pesquisa no campo educacional quanto a sua total eficácia. E isso se deve ao surgimento de outros processos de ensino, que proporcionam algo novo e que demonstram que o grau de satisfação e realização dos estudantes podem ser muito maiores com seu uso.

Diante da observação de inúmeras concepções metodológicas fundamentadas em Aprendizagem Baseada em Problemas (PBL) e Metodologia da Problematização, que também, se configura como estratégia de ensino-aprendizagem, em que os problemas são identificados pelos alunos, pela observação da realidade, houve o interesse em revisitar o referencial teórico relativo à temática, possibilitando contemplar as suas utilidades sob diferentes perspectivas, como: daquele que ensina e daquele que aprende, através de um Mapeamento Sistemático da Literatura (MSL).

2. Problematização e Aprendizagem Baseada em Problemas

Na definição dada por [Delisle e Oliveira 2000](p. 5), PBL é “uma técnica de ensino que educa apresentando aos alunos uma situação que leva a um problema que tem de ser resolvido”. [Lambros 2002] afirma que PBL é um método de ensino que se baseia na utilização de problemas como ponto inicial para adquirir novos conhecimentos.

Portanto, o PBL surge como uma dessas estratégias inovadoras em que os estudantes trabalham com o objetivo de solucionar um problema real ou equivalente a partir de um contexto. Trata-se, portanto, de um método de aprendizagem centrado no estudante, que elimina o papel deste de receptor do conhecimento, para assumir o lugar de protagonista de seu próprio aprendizado. Em conformidade, a Metodologia da Problemática também surge para facilitar a aprendizagem, nela a realidade é problematizada pelos próprios estudantes, enquanto que no PBL, os problemas são cuidadosamente elaborados por uma comissão especialmente designada para este fim.

3. Mapeamento Sistemático da Literatura

Um tipo de revisão complementar à Revisão Sistemática é o de Estudo de Mapeamento Sistemático (também chamado de estudo de escopo). Segundo [Keele et al. 2007], essa é uma “*revisão ampla de estudos primários numa área específica que busca identificar que evidências estão disponíveis nessa área*”. [Petersen et al. 2008] afirmam que “*um mapa sistemático é um método definido de construir um esquema de classificação e estrutura em um campo de interesse*”.

Questões de Pesquisa: Perguntas de pesquisa foram desenvolvidas para que fosse possível dar início ao mapeamento sistemático. O trabalho parte da seguinte questão central: “*Qual o estado da arte da técnica de problematização através do PBL no Ensino Superior?*”. A partir desta, quatro perguntas de pesquisa específicas foram também desenvolvidas: (PP1) Como são construídas as problematizações? (PP2) O que a integração entre PBL e Metodologia da Problemática traz de inovador, quando comparado ao método de ensino tradicional? (PP3) Quais as vantagens da integração entre PBL e Metodologia da Problemática aplicadas ao ensino? (PP4) Quais são as desvantagens da integração entre PBL e Metodologia da Problemática aplicadas ao ensino?.

Estratégia de Pesquisa: Para a realização das buscas dos documentos, foram utilizadas as seguintes bases automáticas: ScienceDirect¹, IEEE Xplore Digital Library² e ACM Digital Library³. De modo a obter todos os documentos relevantes que apresentassem respostas significativas às questões de pesquisa, uma *string* de busca foi construída: (Problematicization OR “troubleshooting Methodology” OR “problem solving methodology” OR “problematicization methodology”) AND (“problem based learning” OR “problem based learned” OR pbl).

Crítérios de Inclusão: (CI1) Estudos que apresentem relação com a questão (e subquestões) de pesquisa; (CI2) Periódicos e artigos publicados em revistas, conferências ou simpósios; (CI3) Estudos que retratam a experiência de PBL dentro do contexto de ensino; (CI4) Pesquisas relacionadas a aplicação PBL- Metodologia da Problemática na educação.

Crítérios de Exclusão: (CE1) Estudos fora do contexto de ensino, educação; (CE2) Estudos que não apresentam aplicação do método PBL dentro do contexto de ensino; (CE3) Documentos incompletos, rascunhos, documentos de compilação dos anais de conferências (*proceeding*), documentos apenas acessíveis por meio da compra e apresentações em slides, documentos que não são artigos; (CE4) Não acessíveis pela

¹<http://www.sciencedirect.com>

²<https://ieeexplore.ieee.org/Xplore/home.jsp>

³<http://dl.acm.org>

Internet; (CE5) Documentos escritos em outras línguas com exceção do inglês; (CE6) Documentos duplicados; (CE7) Documentos que não foram publicados entre os anos de 1997 e 2017.

4. Resultados Parciais e Análise

Esta seção descreve a principal contribuição deste trabalho, a aplicação das técnicas de Mapeamento Sistemático durante a fase de condução que é apresentada conforme protocolo descrito na seção anterior. Entre os meses de Maio e Outubro de 2017, foram realizadas buscas experimentais com base nos critérios de seleção (Inclusão e Exclusão) anteriormente evidenciados. Até o momento, 44 artigos já passaram pela primeira etapa de avaliação (análise pelo título e resumo), onde apenas 37 foram pré-selecionados para a segunda etapa (análise pela introdução). Os próximos pré-selecionados seguirão para a última etapa (Análise completa). A revisão segue em processo de finalização da segunda etapa. Os resultados obtidos em cada base são demonstrados na tabela 1.

Tabela 1. Dados obtidos das Buscas e Resultados Pós-Análise

Avaliação	ScienceDirect	IEEE Xplore	ACM DL
Documentos retornados	20	19	5
1. Título e <i>Abstract</i>	20	15	2
2. Introdução (Em andamento)	-	14	2

5. Conclusão e Trabalhos Futuros

Com relação aos desafios para o desenvolvimento do trabalho, o levantamento dos dados foi inicialmente difícil. A falta e a qualidade de determinados dados foram os principais obstáculos. Apesar dos limites e das possibilidades enfrentadas até o momento, há grande expectativa sobre os resultados do mapeamento, que terão importância para o desenvolvimento e aplicação de alternativas metodológicas, que fundamentem um ensino-aprendizagem com novas possibilidades de situações desafiadoras e ativas aos estudantes no ensino.

Referências

- Delisle, R. e Oliveira, V. (2000). *Como realizar a aprendizagem baseada em problemas*.
 Keele, S. et al. (2007). Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report*. EBSE. sn.
 Lambros, A. (2002). *Problem-based learning in K-8 classrooms: A teacher's guide to implementation*. Corwin press.
 Petersen, K., Feldt, R., Mujtaba, S., e Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE*, volume 8, pages 68–77.

Mineração de Dados Educacionais

Thomás Tabosa de Oliveira¹, Roberta A. de A. Fagundes^{1,2}

¹Universidade de Pernambuco (UPE)
Caruaru– PE – Brasil

²Departamento de Engenharia da Computação – Universidade de Pernambuco, R.
Benfica, 455 – Madalena, CEP:50720-001 - Recife – PE - Brasil

thomas.oliveira@upe.br, roberta.fagundes@upe.br

Abstract. *The School Census is an important tool in improving quality in education for the formulation, implementation and monitoring of educational policies and evaluation of education systems. The proposed article aims to extract information and identify the factors that influence school dropout at the fundamental level of schools in the Northeast region. For data manipulation, the R platform will be the main tool used for data analysis, information extraction and evaluation of the models used in the data mining phase using the CRISP-DM methodology. Currently, this work is in the development phase.*

Resumo. *O censo Escolar é instrumento importante na melhoria da qualidade na educação para a formulação, implementação e monitoramento das políticas educacionais e avaliação dos sistemas de ensino. O artigo proposto tem como objetivo extrair informações e identificar os fatores que influenciam a evasão escolar no nível fundamental das escolas da região Nordeste. Para a manipulação dos dados, a plataforma R será a principal ferramenta utilizada para análise dos dados e extração das informações e avaliação dos modelos utilizados na fase de mineração de dados utilizando a metodologia CRISP-DM.*

1. Introdução

O Censo Escolar, de âmbito nacional, realiza o levantamento de informações estatístico educacionais relativas à Educação Básica, em seus diferentes níveis (educação infantil, ensino fundamental e ensino médio) e modalidades (ensino regular, educação especial e educação de jovens e adultos). O levantamento é feito junto a todos os estabelecimentos de ensino, das redes pública e particular, através do preenchimento de questionário padronizado. Por intermédio do Censo Escolar, o Inep atualiza anualmente o Cadastro Nacional de Escolas e as informações referentes à matrícula, ao movimento e ao rendimento dos alunos, incluindo dados sobre sexo, turnos, turmas, séries e períodos, condições físicas dos prédios escolares e equipamentos existentes, além de informações sobre o pessoal técnico e administrativo e pessoal docente, por nível de atuação e grau de formação.

Pesquisas feitas em vários países mostram que a evasão escolar [Silva 2009] é um problema universal que envolve, para sua solução, distintas intervenções de alunos, pais, professores, autoridades educacionais, formuladores de políticas públicas e lideranças políticas, empresariais e comunitárias. Não obstante as particularidades institucionais de cada nação, a evasão escolar afeta países pobres e ricos, indistintamente, porém, com menor ou maior grau. Nesse sentido, esforços têm sido empreendidos para identificar os indivíduos mais suscetíveis a evadir, seja através da análise das suas características pessoais e familiares, ou pela verificação do seu histórico escolar ou de fatores relacionados às escolas nas quais estão matriculados.

2. Objetivo

Tendo em vista os benefícios sociais e econômicos decorrentes de melhorias educacionais, identificar as causas dos problemas de evasão escolar tem persistido na agenda de pesquisadores e formuladores de políticas sociais. Nesse sentido, este estudo objetiva identificar os fatores determinantes da evasão escolar nas escolas públicas do nível fundamental do Nordeste através dos dados do Censo Escolar [Inep 2017].

Portanto, a presente pesquisa avançará no estudo de mineração de dados através de métodos de *machine learning*, abrangendo aspectos teóricos e práticos, a fim de fornecer subsídios para extração de informações úteis. Para isso, os seguintes objetivos específicos são necessários: Estudar e desenvolver métodos de *machine learning* para mineração de dados utilizando a base de dados do Censo Escolar. Com isso, identificar os fatores que influenciam a evasão escolar do ensino fundamental nas escolas públicas do Nordeste.

3. Metodologia

A metodologia mais aceita em mineração de dados é conhecida por CRISP-DM [Reinartz et al. 1998], sigla do inglês *Cross-Industry Standard Process for Data Mining*. O uso da metodologia CRISP nos permite atacar o problema de extração do conhecimento de forma progressiva e organizada, partindo de uma análise de alto nível, que busca a compreensão do negócio, e indo em direção à definição e implantação de modelos que permitam efetivamente atingir os objetivos da mineração.

Essa metodologia é constituída de seis etapas: Entendimento do negócio, Entendimento dos dados, Preparação dos dados, Modelagem, Avaliação e Aplicação. Cada etapa é subdividida em tarefas definindo assim um processo de trabalho iterativo, ou seja, realizado em ciclos. Dessa forma, um artigo de mineração vai repetir o ciclo de etapas quantas vezes forem necessárias até que os objetivos da mineração tenham sido atingidos.

Nesse artigo é proposto a utilização dessa metodologia no contexto de negócio do Censo Escolar com objetivo extrair conhecimentos úteis para identificar os fatores que influenciam a evasão escolar nas escolas públicas no nível fundamental do Nordeste através do controle direto de grandes bases de dados. Para o desenvolvimento dessa proposta a plataforma R será a principal ferramenta utilizada durante o estudo para criação de um ambiente experimental para extração de informações uteis e avaliação dos modelos propostos para a mineração de dados. No contexto prático será utilizado a base de dados reais do Censo Escolar [Inep 2017] afim de identificar quais fatores influenciam a evasão escolar nas escolas públicas no nível fundamental do Nordeste. O ambiente experimental será destinado para uso desses dados e o desempenho do modelo

será medido pelo erro médio relativo estimado pelo método Monte Carlo [Chernick 1999]. Por fim, serão licitados os fatores que influenciam essa evasão, indicando qual o percentual de influência de cada um para o problema em questão baseado nos dados do Censo Escolar [Inep 2017].

4. Fases de Desenvolvimento

A base de dados do Censo Escolar foi baixada junto com a base da Taxa de Rendimento, onde encontra-se a taxa de abandono das escolas. Foi realizada a filtragem dos dados utilizando o software R para obtenção apenas dos dados da região Nordeste, após foi realizado a remoção das variáveis nulas e as que não seriam úteis para análise do problema, observado por meio da correlação da taxa de abandono com essas variáveis. A base de Matrícula foi filtrada e removidas variáveis nulas por meio do software Knime, pois a mesma se encontrava muito grande para ser processada pelo R e pelo software Knime o processo poderia ser mais prático e vantajoso. O resto do processo foi realizado pelo software R. Por fim todos os dados que são úteis foram guardados na mesma base de dados para posteriormente ser feita a mineração.

5. Conclusão

Este artigo apresentou o objetivo e a metodologia CRISP-DM que será realizada para análise dos fatores que afetam a taxa de abandono. Até o momento já foi realizado a filtragem dos dados e obtenção dos dados principais. As variáveis escolhidas foram aquelas com maiores coeficientes de correlação com a taxa de abandono das escolas. As variáveis estão relacionadas com a escola, se a mesma possui merenda, tempo de duração das aulas e o tipo de dependência (se é estadual, municipal, federal ou particular).

Referências

- Bourdieu, Pierre (1998). A escola conservadora: as desigualdades frente à escola e à cultura. In: NOGUEIRA, Maria Alice; CATANI. Afrânio (Org.). Escritos de educação. Petrópolis, Vozes, p. 40-64.
- Chernick, M. R. 1999. Bootstrap Methods:
- IBGE (2017). Portal do IBGE. Disponível em: <http://www.ibge.gov.br> acessado em 24/02/2017 as 14:33 hs.
- Inep (201). Portal do Inep. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-> acessar acessado em 10/09/2016 as 15:00 hs.
- Han, J. e Kamber M. (2006). Data Mining: Concepts and techniques. Morgan Kaufman Publishers, San Francisco, CA.
- Oreopoulos, Philip (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. Journal of Public Economics, v. 91, n. 11-12, p. 2213-2229.
- Reinartz, T.; Wirth, R.; Clinton, J.; Khabaza, T.; Hejlesen, J.; Chapaman, P.; Kerber, R. (1998). The current CRISP-DM process model for data mining. Berlin: Technical University.

Monitoria como suporte à metodologia de Ensino Baseada em Problemas (PBL) do componente curricular Inglês Técnico

Pedro Victor G. Silva¹, Wylliams B. Santos^{1,2}

¹Universidade de Pernambuco (UPE)
Caruaru– PE – Brasil

²Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Recife – PE - Brasil

pedrovictor07@gmail.com

wbs@upe.br

Resumo. *O objetivo deste artigo é relatar como a disciplina de Inglês Técnico está avançando perante a integração de um ambiente EAD que era inexistente. Esse projeto vem sendo aplicado na Universidade de Pernambuco (Campus Caruaru). Os objetivos são reforçar o canal AcervoUPE no site YouTube, aumentando a biblioteca de informações disponíveis publicamente online, colaborar com o aluno, fazendo acompanhamentos semanais, tudo isso utilizando o método de ensino PBL (Problem Based Learning).*

Abstract. *The objective of this paper is to report how the subject of Technic English is advancing due to E-learning that was inexistent. This project is happening on Universidade de Pernambuco (Campus Caruaru). The objectives are strengthen the AcervoUPE channel on YouTube, increasing the information available publicly online, collaborate with the student, via weekly monitoring all while using the PBL (Problem Based Learning) method.*

1. Introdução

Por meio da tecnologia, estudantes de todo mundo podem, a partir de qualquer local, e a qualquer momento, ganhar conhecimento. Com isso, foi adotado o EAD para a realização desse projeto, por meio de interação com os alunos nas redes sociais e postagem de vídeo aulas no canal AcervoUPE no site YouTube. O método PBL também foi escolhido como abordagem nesse projeto, devido à sua alta taxa de sucesso quando aplicado.

2. Fundamentação

Temos vantagens ao utilizar o EAD, como: Comodidade para o aluno, já que não terá que se locomover para o local de estudos para sanar dúvidas; o conteúdo está disponível 24 horas o que gera flexibilidade de horário para o estudante e interação aluno-monitor a qualquer horário por meio de canais de comunicação indicados pelo monitor.

Aplicando o método PBL, em que o ensinamento é discutido em grupos de aproximadamente oito a dez estudantes para que o conhecimento seja construído com base em casos reais¹ temos uma maior interação aluno-aluno fazendo com que a habilidade de cooperação do aluno melhore.

Também no método PBL, temos vantagens como: Interdisciplinaridade, onde será apresentado para o aluno um conhecimento fragmentado, muitas vezes visto em situações reais e que envolvam vários âmbitos do conhecimento; Melhora da habilidade de comunicação do aluno, visto que trabalhos em grupo pequenos grupos são valorizados; O estudante aprenderá não só o conteúdo abordado na ementa, com a ajuda do professor e monitor, pesquisas serão realizadas pelo aluno, para que resolvam problemas apresentados em sala.

Os alunos foram submetidos à exercícios semanais e encorajados a assistir as videoaulas semanais, que ainda estão sendo gravadas até a presente data.

Na Aprendizagem Baseada em Problemas, o estudo se dá essencialmente na biblioteca, quando os alunos buscam atingir os objetivos cognitivos que elaboraram para alcançar², nesse projeto, EAD também é aplicado, fazendo com que o aluno tenha uma grande biblioteca virtual, além do apoio do monitor o dia inteiro, onde esse problema de estudo essencialmente na biblioteca seja sanado, tendo em vista que o aluno sente-se mais à vontade utilizando a tecnologia ao seu favor, além dele estar perdendo menos tempo na locomoção diária, tendo em vista que grande parte dos alunos não residem na cidade em que estudam.

3. Metodologia

Ao gravar as videoaulas, o assunto abordado da semana é pesquisado, onde os melhores exemplos são selecionados, um script do que é necessário ser falado é produzido, a elaboração dos slides é feita e com a ajuda de um software de gravação de vídeo, as aulas são, por fim, gravadas.

Redes sociais como Facebook e WhatsApp foram essenciais para a implementação. O contato direto com o aluno foi difundido de modo bem mais rápido. Também foram nessas redes sociais onde o aviso de nova videoaula no AcervoUPE e nova atividade semanal foi realizado. Grupos nessas redes sociais foram feitos para que fosse atingida a rápida disseminação de informação.

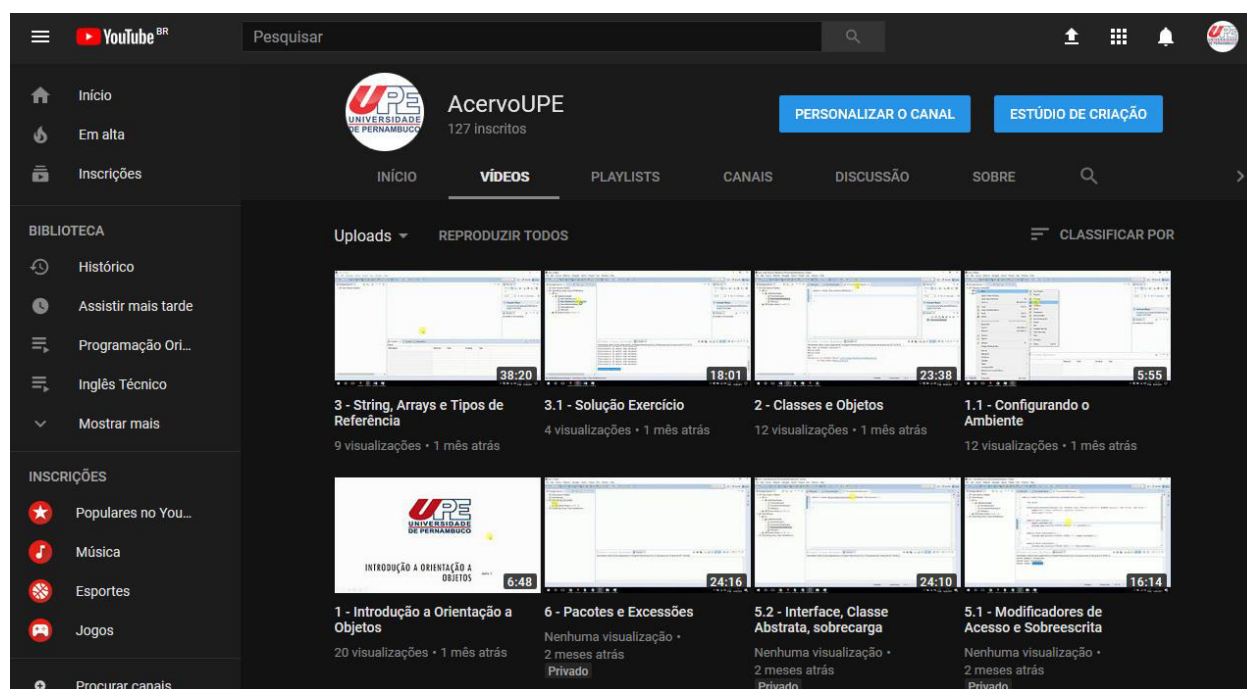


Figura 1. Página inicial do AcervoUPE

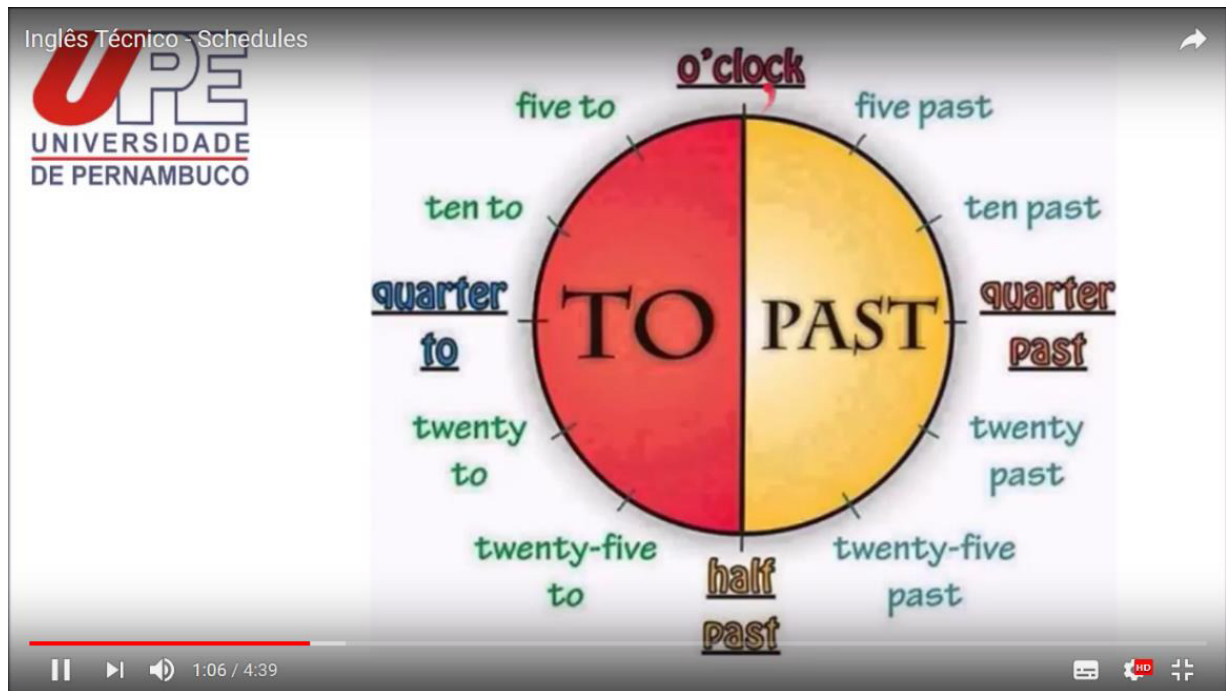


Figura 2. Exemplo de videoaula

4. Resultados

Monitorias presenciais foram marcadas com os alunos, porém não puderam ser realizadas normalmente devido à ausência total de alunos na grande maioria dos dias.

Já o engajamento da turma no ambiente online foi muito satisfatório bem como a quantidade de alunos que realizavam os exercícios propostos semanalmente. Com uma grande taxa de visualização em todos os casos.

O feedback total foi positivo, onde alunos até mesmo realizaram elogios ao modelo das aulas, em especial ao mostrado na figura 2. Nenhum problema ou crítica foi feita.

5. Próximos Passos

As videoaulas continuarão sendo produzidas e daqui para término do projeto serão disponibilizadas as taxas de visualização das videoaulas.

Referências

¹ Perrenoud P.H. (2002) A Prática Reflexiva no Ofício de Professor: Profissionalização e razão pedagógicas, Artmed Editora.

² Berbel N.N. (1998) A problematização e a aprendizagem baseada em problemas: diferentes termos ou diferentes caminhos?, Interface - Comunicação, Saúde, Educação