**Desenvolvimento de um Componente de Suporte à Learning Analytics para um Sistema de Recomendação Híbrido de Trabalhos de Conclusão de Curso**

**1 – INTRODUÇÃO**

Com o aumento do volume de dados gerados a partir da interação de usuários com sistemas on-line, cria-se a oportunidade de efetuar análises mais precisas e aprofundadas desses ambientes, possibilitando a identificação de padrões de relacionamento permitindo a extração de informações implı́citas (JUNIOR; OLIVEIRA, 2016).

A área de Learning Analytics(LA) vem ganhando visibilidade nos últimos por possuir um foco especial na análise deste grande volume de dados, direcionando os resultados especificamente para obter melhorias nos processos dinâmicos de ensino-aprendizagem (MOISSA; KEMCZINSKI; GASPARINI, 2014).

De acordo com (CLOW, 2012), o processo de LA é um ciclo divido em três etapas, onde a primeira dela consiste na coleta de dados, seguida da análise desses dados e, por último, a realização de intervenções.

O processo de LA descrito por (CHATTI et al., 2012) é similar ao descrito por (CLOW, 2012) e o mesmo descreve quatro elementos principais no escopo deste processo: (1) quais dados serão analisados; (2) as pessoas envolvidas nas diferentes etapas do processo; (3) as técnicas a serem aplicadas na análise dos dados; e (4) os objetivos. (CHATTI et al., 2012) categorizou os objetivos de LA como Monitoramento e Análise (MeA); Predição e Intervenção (PeI); Tutoria e Monitoria (TeM); Avaliação e Feedback (AeF); Adaptação (A); Personalização e Recomendação (PeR); e Reflexão (R).

(MOISSA; KEMCZINSKI; GASPARINI, 2014), ao fazer um mapeamento sistemático sobre LA, observou que uma quantidade expressiva tiveram como objetivo central MeA, AeF e PeI. Estes trabalhos utilizaram, principalmente, ferramentas de visualização de informações, estatísticas descritivas, métricas, indicadores, Social Networks Analysis e clusterização. A grande maioria destas análises foram feitas sem o emprego de um algoritmo específico, exceto pelo k-means para clusterização. Ao fazer um levantamento quando os instrumentos utilizados, destacam-se o MS Excel, Moodle, linguagem de programação R, MySQL e Google Docs. Por fim, as maiores contribuições proporcionadas por estes trabalhos foram implementação de ferramentas, conhecimento sobre o processo de ensino-aprendizagem, desenvolvimento de novas abordagens e produção de frameworks para LA.

Análise multivariada de dados é um ramo da análise de dados que visa o estudo do relacionamento ou conjunto de relações entre múltiplas variáveis, possibilitando o levantamento de questões específica e com alto nível de complexidade (HAIR et al., 2009)

Visualização de informações, ou visualização de dados, é uma área de consiste na aplicação de diferentes técnicas, geralmente interativas, que tem como objetivo principal auxiliar no processo de análise de dados, seja ela univariada ou multivariada, através de representações gráficas (FREITAS et al., 2001).

Como ferramenta produzida pelos Projetos de Apoio à Iniciação Cientı́fica nas edições 2015/2016 e 2016/2017, dos alunos Lucas Gabriel Coimbra Evangelista (código no Sisproj 33958 e 43715, respectivamente) e Salomão Augusto Vilaça Corrêa (código no Sisproj 33662, edição 2015/2016), orientados pelo Prof. Dr. Fábio Santos da Silva, o Tenebris é um sistema de recomendação para artigos acadêmicos que usa filtragem hı́brida e técnicas de aprendizagem de máquina, implementado na UEA/EST.

No escopo desse projeto será considerado o desenvolvimento de suporte à LA para a ferramenta Tenebris, (1) analisando dados provenientes desta plataforma, (2) gerados por alunos e professores da UEA/EST, (3) utilizando diferentes técnicas de análise e visualização de dados, bem como técnicas de aprendizagem de máquina, (4) para monitorar o desenvolvimento de pesquisas científicas (MeA), auxiliar os professores nos processos de TeM (seja para trabalhos de conclusão de curso, iniciação cientı́fica, iniciação à docência, etc.), AeF para a universidade em relação as pesquisas desenvolvidas e, por fim, fazer predições em relação as áreas de pesquisas mais promissoras para a universidade.

Para apresentar o projeto de iniciação científica que é proposto neste documento, as seções a seguir detalham os demais componentes deste projeto. A justificativa para o desenvolvimento do mesmo é apresentada na Seção 2. Os objetivos a serem alcançados são expostos na Seção 3. A Metodologia a ser seguida é mostrada na Seção 4, cujo cronograma de execução é mostrado na Seção 5.

**2 – JUSTIFICATIVA**

O desenvolvimento de um componente de suporte à LA para a ferramenta Tenebris é importante por diversas razões.

Em termos de análise e visualização de dados, tem-se que o aluno irá lidar com diversos modelos estatísticos e diferentes técnicas de visualização de dados, que proporcionam suporte ao desenvolvimento de várias áreas do conhecimento, a exemplo da Biologia, Geografia, Matemática, Engenharia, dentre outras. Ao implementar tais métodos e técnicas, aumenta-se o conhecimento sobre as variadas aplicações destes, bem como proporciona-se embasamento teórico e prático para utilização destas técnicas em trabalhos futuros do próprio, a exemplo do trabalho de conclusão de curso.

No que se refere a implementação de LA, tem-se um favorecimento no seu uso para análise e visualização de dados voltados para a melhoria do processo de aprendizagem, bem como para predição de eventos, possibilitando a realização de intervenções para evitar problemas, como a queda do rendimento acadêmico e científico dos alunos e professores. Tendo em vista que a área de LA é consideravelmente recente e não há nenhuma ferramenta desta área implementa na UEA/EST, este projeto pode ser uma valorosa contribuição para a universidade, ajudando na melhoria do nı́vel das produções científicas, aumento da quantidade de artigos publicados e melhor direcionamento de recursos para pesquisa e desenvolvimento.

No tocante ao uso da ferramenta Tenebris, tem-se um favorecimento no seu uso para obtenção de dados de interações entre usuários e sistemas on-line. A escolha do Tenebris deu-se por haver o livre acesso ao código fonte da ferramenta e conhecimento do seu processo de desenvolvimento, facilitando a integração do componente de suporte à LA.

Em se tratando do desenvolvimento de habilidades e competências do aluno que irá participar deste projeto, é importante ressaltar que há uma forte relação com o perfil do egresso de Engenharia da Computação, pois há a análise e visualização de dados; desenvolvimento de componentes a serem acoplados a sistemas já em funcionamento; e, por fim, a análise comparativa entre produtos produzidos. Todas estas atividades colaboram para um senso crítico do aluno em termos do conhecimento com que ele lida e produz.

**3 – OBJETIVO**

O objetivo geral do projeto proposto consiste em desenvolver, especificamente, um componente de suporte à Learning Analytics para sua incorporação em um sistema de recomendação hı́brido de artigos, dissertações, teses e trabalhos conclusão de curso. O sistema visa prover aos alunos e professores acesso personalizado aos dados gerados por pesquisas e publicações de trabalhos acadêmicos oriundos da EST/UEA. Para alcançar esta meta, alguns objetivos específicos precisam ser contemplados, a citar:

1. Formular um referencial teórico que caracterize os principais aspectos de Learning Analytics;
2. Levantar trabalhos relacionados empregados a unidades de ensino superior;
3. Identificar e avaliar ferramentas que deem suporte à aplicação de Learning Analytics;
4. Projetar e implementar um protótipo de Learning Analytics para análise e visualização de dados;
5. Avaliar, através de testes sugeridos pela literatura, a qualidade das informações geradas pelo protótipo produzido.

**4 – METODOLOGIA**

A metodologia para o desenvolvimento deste projeto de iniciação cientı́fica consiste, inicialmente, no estudo dos conceitos relacionados a Learning Analytics, que proporcionarão ao aluno o ferramental teórico necessário para compreendê-la do ponto de vista educacional e também do prático, por meio do levantamento de aplicações da mesma. Ao final desta etapa, o aluno irá construir uma base de dados estática para ser utilizado na primeira fase do projeto.

O passo seguinte é o estudo de técnicas de análise e visualização de dados, fazendo um levantamento das ferramentas e linguagens de programação mais apropriadas para implementação destas técnicas e diferentes métodos a serem utilizados, baseando-se nas recomendações sugeridas pela literatura para tratar os tipos de dados coletados. Ao final desta etapa, espera-se que o aluno tenha feito a análise dos dados coletados e produção de gráficos e tabelas para melhor visualização dos dados.

O próximo passo é o desenvolvimento de uma ferramenta de aprendizagem de máquina, tendo em vista a realização projeções das áreas de pesquisas mais promissoras para UEA/EST, usando como base os anais dos principais eventos nacionais relacionados as principais áreas de pesquisa da universidade.

Após a implementação das fases anteriores, será realizada a análise de qualidade da visualização de informações, baseada em questionários respondidos por usuários e métricas sugeridas pela literatura, e da ferramenta de aprendizagem de máquina.

Após os testes, será feito acoplamento do protótipo ao Tenebris, onde será realizada uma nova bateria de testes, estes agora on-line, e a adaptação dos algoritmos de visualização de dados e de predição, para que eles funcionem de forma dinâmica, sendo atualizado conforme novos dados sejam gerados pelas interações.

Ao final de cada uma das etapas de desenvolvimento, o aluno irá disponibilizar o programa produzido e a documentação resultante em um repositório de software na internet. Este será o meio de divulgação dos resultados produzidos.

Além das atividades relatadas, não se pode deixar de mencionar a produção dos relatórios parcial e final, que são essenciais para o andamento do projeto. Esta atividade possuirá um acompanhamento mais minucioso e utilizará artefatos produzidos ao longo do andamento do projeto para a composição destes relatórios.

**5 – CRONOGRAMA**

Uma visão geral do cronograma das atividades a serem desenvolvidas ao longo deste projetopodem ser vistas na Tabela 1. Elas possuem relação com as atividades listadas na Seção 4, a qual detalha a metodologia que o projeto de pesquisa deverá seguir.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2017 | | | | | 2018 | | | | | | |
|  | 08 | 09 | 10 | 11 | 12 | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| Estudo dos Conceitos Relacionados a Learning  Analytics | X |  |  |  |  |  |  |  |  |  |  |  |
| Estudo de Técnicas de Análise e Visualização de  Dados |  | X |  |  |  |  |  |  |  |  |  |  |
| Implementação da Ferramenta de Análise e  Visualização de Dados |  |  | X | X |  |  |  |  |  |  |  |  |
| Implementação da Ferramenta de Aprendizagem  de Máquina |  |  |  |  | X | X | X |  |  |  |  |  |
| Acoplamento do Protótipo ao Tennebris |  |  |  |  |  |  |  | X | X | X |  |  |
| Análise de Qualidade |  |  |  |  |  | X |  |  | X |  | X |  |
| Elaboração de Relatórios e Artigos | X | X | X | X | X | X | X | X | X | X | X | X |

Tabela 1: Cronograma de atividades levando em consideração os doze meses (08/2017 a

07/2018) para a realização do PAIC.

**6 – MATERIAIS**

01 – HD externo para armazenamento e backup;

02 – Papel ofício (resma);

03 – cartucho para impressora;

**REFERÊNCIAS**

CHATTI, M. A. et al. A reference model for learning analytics. International Journal of Technology Enhanced Learning, Inderscience Publishers, v. 4, n. 5-6, p. 318–331, 2012.

CLOW, D. The learning analytics cycle: closing the loop effectively. In: ACM. Proceedingsof the 2nd international conference on learning analytics and knowledge. [S.l.], 2012. p.134–138.

FREITAS, C. M. D. S. et al. Introdução à visualização de informações. Revista de informáticateórica e aplicada. Porto Alegre. Vol. 8, n. 2 (out. 2001), p. 143-158, 2001.

HAIR, J. F. et al. Análise multivariada de dados. [S.l.]: Bookman Editora, 2009.

JUNIOR, C. B. da S.; OLIVEIRA, I. C. A. de. LEARNING ANALYTICS: REVISÃO DA LITERATURA E O ESTADO DA ARTE. 2016. hhttp://www.abed.org.br/congresso2016/trabalhos/329.pdfi. Acessado em 30 de maio de 2017.

MOISSA, B.; KEMCZINSKI, A.; GASPARINI, I. Learning analytics: entendendo e otimizando o aprendizado. In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. Companion Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems. [S.l.], 2014. p. 59–60.