

**UNIVERSIDADE DE TAUBATÉ**  
**Pedro Luiz da Silva Pereira**

**APLICAÇÃO DAS FERRAMENTAS DE MINERAÇÃO DE DADOS E  
BUSINESS INTELLIGENCE NA PLATAFORMA MOODLE**

**Taubaté – SP**  
**2019**

**Sistema Integrado de Bibliotecas – SIBi / UNITAU**  
**Biblioteca Setorial do Departamento de Informática**

P436a     Pereira, Pedro Luiz da Silva  
              Aplicação das ferramentas de Mineração de dados e *Business Intelligence* na plataforma *Moodle* / Pedro Luiz da Silva Pereira. - 2019.  
              47f. :il.

              Monografia (especialização) - Universidade de Taubaté,  
              Departamento de Informática, 2019.

              Orientação: Prof. Me. Dawilmar Guimarães de Araujo,  
              Departamento de Informática.

              1. Administração de projetos - Gestão. 2. *Business Intelligence*.  
              3. Mineração de dados (Computação). I. Título. II. Universidade de  
              Taubaté.

CDD 658.404

**Pedro Luiz da Silva Pereira**

# **APLICAÇÃO DAS FERRAMENTAS DE MINERAÇÃO DE DADOS E BUSINESS INTELLIGENCE NA PLATAFORMA MOODLE**

Monografia, apresentada para obtenção do Certificado de Especialização através do curso de Pós-Graduação em Gerenciamento de Projetos com ênfase em Business Intelligence na Universidade de Taubaté - UNITAU

Orientador: Prof. Me. Dawilmar Guimarães Araújo

**Taubaté – SP  
2019**

**Pedro Luiz da Silva Pereira**

## **APLICAÇÃO DAS FERRAMENTAS DE MINERAÇÃO DE DADOS E BUSINESS INTELLIGENCE NA PLATAFORMA MOODLE**

Trabalho de Conclusão de Curso aprovado pela Banca Examinadora da Monografia, apresentada para obtenção do Certificado de Especialização através do curso de Pós-Graduação em Gerenciamento de Projetos com ênfase em Business Intelligence na Universidade, com Linha de Pesquisa bibliográfica.

Data: \_\_26\_\_ / \_\_10\_\_ / \_\_2019\_\_ Resultado: \_\_Aprovado\_\_(9,5)\_\_\_\_\_

### **BANCA EXAMINADORA**

Prof. Me. Dawilmar Guimarães Araújo

Universidade de Taubaté

Assinatura\_\_\_\_\_

Prof. Dr. José Carlos Lombardi

Universidade de Taubaté

Assinatura\_\_\_\_\_

Profa. Dra. Rita de Cássia Rigotti Vilela Monteiro

Universidade de Taubaté

Assinatura\_\_\_\_\_

**Taubaté –SP  
2019**

Minha família pela compreensão e incentivo

## AGRADECIMENTOS

Ao Prof. Dr. Luis Fernando de Almeida pelo incentivo para realização da Pós-Graduação.

Ao Prof. Me. Dawilmar Guimarães Araújo na orientação da elaboração da monografia.

Ao Prof. Me. Fernando Gama nas referências técnicas sobre ETL e Big Data.

Ao Prof. Me. Antônio Ricardo Mendrot nas referências de gestão de projetos, Data warehouse e ferramentas de BI.

A Profa. Dra. Rita de Cássia Rigotti Vilela Monteiro na condução e revisão da monografia.

Aos colegas de turma os quais minha gratidão às contribuições de agregação ao conhecimento adquirido.

A Instituição UNITAU, que me acolheu como aluno em toda minha trajetória acadêmica.

“A única maneira de fazer um bom trabalho é amando o que você faz. Se você ainda não encontrou, continue procurando. Não se desespere. Assim como no amor, você saberá quando tiver encontrado.”

Steve Jobs

## RESUMO

O Moodle é utilizado por instituições de ensino e empresas em mais de 227 países do mundo, o Brasil atualmente é o 4º país no *ranking* do uso plataforma com 4.344 registros de uso da plataforma (Moodle 14/09/2019, <https://moodle.org/>). Atualmente a adoção do Moodle como plataforma para fazer a gestão do ensino a distância que possui licenciamento de código aberto, que são uns dos diferenciais no momento da escolha e adoção, porém após a adoção apesar de estar disponibilizado nem todos os recursos são utilizados, dessa forma, podemos fomentar os usuários oferecendo oportunidades no crescimento do uso da plataforma? Alcançar o diagnóstico da base de dados da plataforma de ensino Moodle a fim de identificar oportunidades de melhorias na experiência da jornada de professores, alunos e gestores de educação do uso dentro da plataforma. Realizar uma análise sobre a base de dados da plataforma Moodle em 3 momentos: Identificar os módulos não utilizados e oferecer um relatório do que ele pode ser conquistado utilizando a ferramenta. Aplicar as técnicas da mineração dos dados, aplicar regras de clusterização e associação, onde o resultado apoiará o embasamento das informações. Gerar *dashboard* dos cursos e itens relacionados ao curso com informações relevantes para apoiar a tomada de decisão e professores e gestores de educação. A oportunidade de criar uma solução a ser reutilizada em outras instituições inclusive no futuro, além do caso de uso em questão, como uma ferramenta de apoio ao uso de uma ferramenta que hoje tem números que apoiam o crescimento do ensino a distância, o que tem sido uma oportunidade de levar conhecimento a diversas pessoas ao redor do mundo.

Palavras-chave: Gerenciamento de Projetos; Plataforma Moodle; Business Intelligence; Mineração de Dados;

## ABSTRACT

Moodle is used by educational institutions and companies in over 227 countries worldwide, Brazil is currently the 6th country in the platform usage ranking with 4,344 platform usage records (Moodle 14/09/2019, [https://moodle.org /](https://moodle.org/)). Currently, the adoption of Moodle as a platform to manage distance learning that has open source licensing, which are one of the differentials at the time of choice and adoption, but after adoption, although it is available, not all resources are used. So can we nurture users by offering opportunities to grow platform usage? Achieve diagnostics of the Moodle learning platform database to identify opportunities for improvement in the journey experience of teachers, students and education managers of use within the platform. Conduct an analysis of the Moodle platform database in 3 moments: Identify unused modules and provide a report of what they can achieve using the tool. Apply data mining techniques, apply clustering and association rules, where the result will support the information base. Generate course dashboard and course related items with relevant information to support decision making and teachers and education managers. The opportunity to create a solution to be reused in other institutions, including in the future, in addition to the use case in question, as a tool to support the use of a tool that today has numbers that support the growth of distance learning, which has been an opportunity to bring knowledge to many people around the world.

Key-words: Project Management; Moodle Platform; Business Intelligence; Data Mining.

## LISTA DE ILUSTRAÇÕES

|   |    |
|---|----|
| FIGURA 01: Processo de Business Intelligence .....  | 19 |
| FIGURA 02: Como funciona o algoritmo Apriori (Traduzido) .....  | 21 |
| FIGURA 03: Como funciona o algoritmo EM (Expectation Maximisation)<br>(Traduzido) .....                           | 22 |
| FIGURA 04: Ciclo de uma Sprint, repetido até o produto e evolução (Traduzido)<br>.....                            | 23 |
| FIGURA 05: Escolha da metodologia Scrum em um projeto (traduzido) .....   | 23 |
| FIGURA 06: Modelo evolutivo usado no processo de engenharia de software ...                                       | 24 |
| FIGURA 07: Arquitetura do banco de dados MySQL .....  | 25 |
| FIGURA 08: Arquitetura de ferramentas e frameworks utilizados na aplicação ...                                    | 27 |
| FIGURA 09: Metodologia SCRUM adotada no projeto (adaptado) .....  | 28 |
| FIGURA 10: Aplicação do Javadoc para geração da documentação técnica .....  | 30 |
| FIGURA 11: Estrutura do processo de ETL no projeto .....  | 32 |
| FIGURA 12: Arquitetura, aplicação e arquitetura utilizando JUnit (tradução).....                                  | 33 |
| FIGURA 13: maven-archetype-webapp é um arquétipo qual gera um exemplo de<br>um projeto webapp Maven .....         | 34 |
| FIGURA 14: Arquitetura Macro da IDE Eclipse (traduzido) .....   | 35 |
| FIGURA 15: Macroprocesso demonstrando o uso do Github .....   | 36 |
| FIGURA 16: Visão geral do funcionamento da solicitação do usuário ao servidor<br>de aplicação e seu retorno ..... | 37 |
| FIGURA 17: Arquitetura LAMP (Linux, Apache, MySQL e PHP) do Moodle .....  | 37 |
| FIGURA 18: Ideia geral da mineração de dados .....  | 38 |
| FIGURA 19: Módulo Analytics .....   | 41 |
| FIGURA 20: Exibição da contabilização dos módulos na ferramenta .....   | 43 |
| FIGURA 21: Exibição da contabilização de informações de módulos não<br>utilizados.....                            | 43 |
| FIGURA 22: Visão geral das etapas para processamento dos relatórios de<br>mineração .....                         | 44 |

## LISTA DE TABELAS

|   |    |
|---|----|
| TABELA 01: Números Relevantes da plataforma Moodle .....  | 14 |
| TABELA 02: Frameworks utilizados no desenvolvimento ..... | 26 |
| TABELA 03: Número de Tabelas por módulo .....             | 40 |

## LISTA DE ABREVIATURAS E SIGLAS

|      |                                       |
|------|---------------------------------------|
| BI   | Business Intelligence                 |
| EM   | Expectation Maximisation              |
| ETL  | Extract Transform Load                |
| IDE  | Integrated Development Environment    |
| LAMP | Linux, Apache, MySQL e PHP            |
| POM  | Project Object Model                  |
| HTTP | Hypertext Transfer Protocol           |
| J2EE | Java Platform, Enterprise Edition     |
| MIT  | Massachusetts Institute of Technology |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 – Introdução</b> .....   | <b>13</b> |
| 1.1 - Objetivo Geral .....  | 13        |
| 1.2 - Objetivos Específicos .....   | 13        |
| 1.3 - Relevância .....  | 14        |
| 1.4 - Delimitações .....  | 15        |
| 1.5 - Justificativa do trabalho .....                                     | 16        |
| 1.6 - Estrutura do Trabalho .....   | 16        |
| <b>2 – Fundamentações Teóricas</b> .....                                  | <b>18</b> |
| 2.1 - <i>Plataforma Moodle – tecnologias de ensino a distância</i> .....  | 18        |
| 2.2 - <i>Business Intelligence</i> .....                                  | 19        |
| 2.3 - <i>Mineração de dados</i> .....                                     | 19        |
| 2.4 - <i>Gestão do Projeto</i> .....                                      | 22        |
| 2.5 - Ciclo de Vida do Software .....                                     | 23        |
| 2.6 - Dashboards .....  | 25        |
| 2.6 - <i>Database</i> .....   | 24        |
| 2.7 - ETL .....   | 25        |
| <b>3 - O desenvolvimento do projeto</b> .....                             | <b>27</b> |
| 3.1 - Gestão de Projetos .....  | 29        |
| 3.10 - Utilizando os recursos de mineração de dados .....                 | 38        |
| 3.11 - <i>Dashboards</i> .....  | 38        |
| 3.2 - ETL .....   | 31        |
| 3.3 - Gerenciador de <i>frameworks</i> de desenvolvimento .....           | 32        |
| 3.4 - Automatizando execução de testes unitários .....                    | 33        |
| 3.5 - Estrutura da solução técnica .....                                  | 33        |
| 3.6 - Configurando um ambiente desenvolvimento integrado .....            | 34        |
| 3.7 - Gerenciamento de código fontes .....                                | 35        |
| 3.8 - Configurando o servidor de aplicação .....                          | 36        |
| 3.9 - Gerenciamento de armazenamento dos dados na plataforma Moodle ..... | 37        |
| <b>4 - Desenvolvimento</b> .....  | <b>39</b> |
| 4.1 - Metodologia de desenvolvimento .....                                | 39        |
| 4.2 - Interação das fases .....   | 39        |
| 4.3 - Processo de contabilização dos módulos .....                        | 41        |
| 4.4 - Processo de processamento de mineração de dados .....               | 42        |
| 4.5 - Apresentação dos resultados para o usuário final .....              | 42        |
| <b>5 - Resultado</b> .....  | <b>43</b> |
| <b>6 - Conclusão</b> .....  | <b>46</b> |
| Referências bibliográficas .....  | 47        |

# 1 – Introdução

O Moodle é uma plataforma de ensino a distância que tem 103.978 instâncias registradas em mais de 227 países, onde o Brasil enquadra-se em 4º lugar no uso da plataforma, segundo o moodle.org (2019).

O licenciamento é de código aberto, ou seja, não há custos de aquisição de licenças para o uso o que também vem a ser parte do fator de decisão de adoção da plataforma em companhias e instituições de ensino em geral.

Às customizações a serem feitas na plataforma ficam sobre a gestão do administrador da ferramenta, bem como o fomento do uso por parte dos professores de como fazer a gestão de conteúdos e avaliações na plataforma.

## 1.1. Objetivo Geral

Diagnosticar a base de dados da plataforma de ensino a distância Moodle para identificar oportunidades de melhorias e crescimento através da experiência da jornada de: professores, alunos e gestores de educação no uso da plataforma.

## 1.2. Objetivos Específicos

Realizar uma análise sobre a base de dados da plataforma Moodle em 3 momentos:

- Identificar os módulos não utilizados e oferecer um relatório do que pode ser conquistado utilizando a ferramenta;
- Aplicar as técnicas de mineração de dados, através das regras de clusterização e associação, onde o resultado da mineração apoiará o embasamento das informações;
- Gerar *dashboards* dos cursos e seus itens relacionados com informações relevantes para apoiar a tomada de decisão de professores e gestores de educação.

### 1.3. Relevância

A importância do nosso estudo é disponibilizar para os administradores da plataforma, educadores e gestores da área de educação uma ferramenta que possa apoiar a oportunidade de explorar a base de dados de uma instalação do Moodle, fomentando o uso dos recursos não utilizados, analisar utilizando ferramentas de mineração de dados, onde resultados serão exibidos para o usuário final para ajudar a tomada de decisão no uso da plataforma.

A oportunidade de criar uma solução a ser reutilizada em outras instituições, além do caso de uso em questão, como uma ferramenta de apoio ao uso da plataforma que faz a gestão do ensino a distância, o que tem sido uma oportunidade de levar a novos conhecimentos a diversas pessoas ao redor do mundo. Conhecimento este, podendo ser um curso livre, ou com relevância acadêmica, ou seja, ensino médio, Graduação, Pós-Graduação. Por ter a opção de instalação em data center local a plataforma pode ser adotada pelas instituições com baixo custo para infraestrutura, uma vez que a plataforma além de muito utilizada no mundo tem licenciamento tipo código aberto. Alguns números relevantes:

TABELA 01: Números Relevantes da plataforma Moodle no mundo

| Indicador         | Valor         |
|-------------------|---------------|
| Sites Registrados | 103.978       |
| Países            | 227           |
| Cursos            | 19.762.084    |
| Usuários          | 168.883.356   |
| Inscrições        | 798.572.063   |
| Fóruns            | 362.975.292   |
| Recursos          | 172.646.003   |
| Pesquisas         | 1.730.458.344 |

fonte: moodle.org (acessado em: 14/09/2019)

## 1.4. Delimitações

Este trabalho será realizado utilizando uma base de uma instância Moodle na versão 2.7 de uma instituição de ensino acadêmica localizada no Vale do Paraíba. A ferramenta que será construída fará uma análise sobre a base a partir de atividades de consulta, retornando o diagnóstico em forma de alertas com informações sobre os percentual dos módulos utilizados, prover informações para mineração de dados nestes casos utilizando o algoritmo *Apriori* para associação e o EM para clusterização, e sobre os módulos que não estão funcionando prover informações relevantes sobre o módulo, a fim de fomentar o uso do módulo na instituição. Sobre as ferramentas e *frameworks* utilizadas, delimita-se o uso nas seguintes versões:

TABELA 02: *Frameworks* utilizados no desenvolvimento

| Framework       | Versão         | Aplicabilidade   |
|-----------------|----------------|--|
| MySQL dump      | 10.13          | Versão do utilitário gerador do backup da base de dados.   |
| MySQL Distrib   | 5.7.23         | Versão da distribuição do Servidor MySQL, utilizado na base de dados.  |
| Linux           | ubuntu 16.04.1 | Versão do sistema operacional onde a instalação do Moodle objeto de estudo foi instanciada.  |
| POM             | 4.0.0          | <i>Project Object Model</i> ferramenta utilizada no projeto Maven Java, contendo os <i>archetypes</i> para os componentes Web e informações de configuração de estrutura e <i>frameworks</i> .   |
| JUNIT           | 5.4.2          | <i>Framework</i> responsável para fazer a gestão de testes unitários, cujo o objetivo é garantir qualidade funcional do código.  |
| Servlet         | 2.5            | Programa na linguagem Java capaz de trabalhar em um servidor de aplicação através do modelo de requisições e respostas de informações definidas através do protocolo HTTP ( <i>Hyper Text Transfer Protocol</i> , comunicação WEB entre cliente e servidor. ).   |
| Weka            | 3.8.2          | Coleção de algoritmos utilizados em aprendizado de máquina, para realização de atividades de mineração de dados.   |
| MySQL Connector | 8.0.16         | Conector Java utilizado na comunicação entre a aplicação Java e o banco de dados MySQL.  |
| Maven           | 2.6            | Permitir a compreensão completa do esforço do desenvolvimento, os principais objetivos são: <ul style="list-style-type: none"><li>- facilitar o processo de construção;</li><li>- construir um sistema uniforme;</li><li>- utiliza às melhores práticas de desenvolvimento;</li><li>- transparência na migração de novas versões de <i>frameworks</i>;</li></ul> |
| Eclipse         | neon.3         | IDE ( <i>Integrated Development Environment</i> ), utilizado para construir efetivamente a aplicação.  |
| Tomcat/Apache   | 8.0/7.07.0     | Servidores de aplicação para uso local para desenvolvimento e remoto pelos usuários finais.  |

|            |   |  |
|------------|---|--|
| GZip       |   | Utilitário para instalação da aplicação no servidor externo.   |
| JDK        | 8   | JDK ( <i>Java Developer Kit</i> ) ambiente para desenvolvimento de aplicações utilizando a linguagem de programação JAVA.  |
| VirtualBox | Version 6.0.10<br>r132072<br>(Qt5.6.3)<br>Sistema Operacional:<br>Ubuntu 16.04<br>LTS | É uma plataforma de virtualização de aplicações. Estende a capacidade do seu computador existir em qualquer sistema operacional, dentro de uma máquina virtual, podendo ser mais um computador ao mesmo tempo. |

fonte: próprio autor

## 1.5. Justificativa do trabalho

Após o ciclo inicial de adoção da plataforma Moodle em uma instituição ou companhia que consiste na instalação do ambiente e determinar como ferramenta de ensino à distância. É necessário que os profissionais nestas organizações que fazem a gestão da ferramenta possam ter fatos e dados que justifiquem a ampliação do serviço, ou ainda um retorno sobre o investimento de: tempo, treinamento, infraestrutura, profissionais focados no resultado da ferramenta, bem como a qualidade de como o conhecimento é difundido, adquirido e compartilhado através do uso da plataforma.

Neste contexto uma ferramenta que analise a base de dados a fim de fomentar o uso de módulos não utilizados, com o intuito de alertar sobre propostas de novos focos junto com seus ganhos na ferramenta, bem como os módulos existentes, ajudar a identificar casos de sucesso para se tornar modelos e também oportunidades de ampliar a satisfação da difusão de conhecimento através do ensino a distância utilizando a ferramenta, em um momento seguinte uma visão geral de como estão às métricas dos conteúdos organizados através de cursos.

## 1.6. Estrutura do Trabalho

O primeiro capítulo trata sobre a introdução. Neste capítulo foi explanado o tema, a importância e a abordagem do estudo. Ainda neste capítulo foram apresentados o objetivo geral e os objetivos específicos, a relevância, justificativa e estrutura da composição deste trabalho.

No segundo capítulo ocorreu a fundamentação teórica, neste momento foi introduzido ao leitor o estado da arte sobre as ciências e conceitos que entornam neste trabalho, nela foi abordada conceitualização e suas divisões de gerenciamento de projetos dando ênfase às ferramentas utilizadas para construção da ferramenta chamada de Moodle *Analytics*.

No terceiro capítulo a abordagem sobre as metodologias utilizadas e aplicadas neste trabalho.

No quarto capítulo o desenvolvimento relatório e *dashboards* para análise das informações da base de dados, objeto deste estudo.

No quinto capítulo foi apresentado os resultados obtidos, produto do desenvolvimento realizado.

No sexto e último capítulo foi realizado a conclusão do trabalho, e o desfecho do raciocínio e consequências desses argumentos explorados.

## 2 – Fundamentações Teóricas

Para construção da ferramenta, foi realizada uma análise técnica a fim de garantir a qualidade do produto, bem como a governança durante a execução do projeto, com este propósito em mãos e com o desafio a alcançar, foi elaborada a arquitetura de ferramentas e *frameworks* do projeto para ser implementado com sucesso.

### 2.1 Plataforma Moodle – tecnologias de ensino a distância

O Moodle é uma plataforma de ensino a distância utilizado em 227 países por diversos tipos de instituições relacionados a ensino variando de um curso livre até uma formação acadêmica em nível de Pós-Graduação. Uma das principais características do Moodle é o fato de possuir licenciamento aberto, ou seja, quando uma instituição o adota como plataforma de ensino a distância a proposta é ter o custo apenas de hospedagem no *data center* local ou servidor de hospedagem. Por se tratar de um código aberto o Moodle possui uma das maiores equipes de desenvolvimento do mundo encorajando pessoas ao redor do mundo a colaborar na evolução da plataforma.

Para os gestores de educação, educadores e alunos, o Moodle oferece um ambiente seguro e de forma integrada, garantindo robustez e personalização a sua necessidade de plataforma de ensino a distância em instituições acadêmicas ou no segmento corporativo.

O Moodle possui uma interface simplificada com a finalidade de implementar melhorias na experiência de uso, tem um projeto global de atualização de versão atendendo necessidades globais reportadas pelos membros da comunidades, tem a facilidade de personalizar o idioma da ferramenta, totalmente flexível para customizações completas do ambiente, permite também uma escalabilidade de infraestrutura de modo a manter operante a instância, pode ser utilizado de qualquer lugar e em qualquer dispositivo, possui também recursos adicionais que podem ser baixados e instalados.

## 2.2 Business Intelligence

O conceito BI (*Business Intelligence*) está relacionado a tecnologias, práticas e aplicações que de forma integrada promove a coleta de informações e a integração entre elas, de forma a facilitar a análise e apresentação de informações de negócio, cujo principal objetivo é de suportar às organizações auxiliando na melhor e mais rápida tomada de decisão.

A importância de ferramentas de BI é prover informações históricas, atualizadas e preditivas referente aos processos de negócio da organização.

A figura 01 abaixo demonstra o processo de BI desde sua origem das informações, geralmente os banco de dados dos sistemas nas organizações, mas podem ter serviços web e arquivos, onde através do processo de ETL (Extração, Transformação e Carga) os dados são tratados e vão para os armazéns de dados (*data warehouse*), onde a base para os dashboards e outros serviços web são disponibilizados através dos cubos de dados.

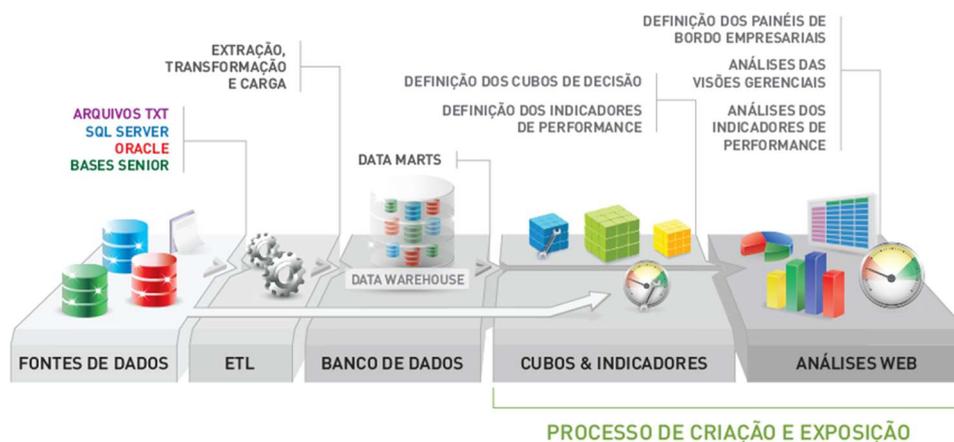


FIGURA 01: Processo de *Business Intelligence*  
fonte: <https://senior.com.br>, acessado em 20/10/2019

## 2.3 Mineração de dados

Segundo Alpaydin (2014), estamos na era do “*Big Data*”. A algum tempo atrás somente às companhias continham dados. Este usados em centros de computação onde estes dados eram armazenados e processados. Primeiro com a chegada do computador pessoal e depois com uso dos computadores em rede e

internet, todos nós começamos a produzir dados. A todo tempo nós compramos um produto, alugamos um filme, visitamos uma página web, escrevemos um blog ou postamos em uma rede social e ainda quando apenas caminhamos ou dirigimos, nós continuamos gerando dados.

Cada indivíduo não apenas consome dados, mas também produz dados, seja qual for a localização é possível consumir produtos e serviços, onde o provedor deve atender os pedidos com base nas necessidades e interesses preditivos.

Para resolver um problema computacional é necessário um algoritmo. Um algoritmo é uma sequência de instruções que devem ser carregadas como entrada que após transformada é gerada uma saída. No entanto para algumas atividades não temos um algoritmo. Não é possível identificar o processo completamente, mas acreditar de que é possível construir uma boa e usual aproximação. Esta aproximação não abrange tudo, mas talvez ainda contabilizar uma parte dos dados. Nós ainda podemos detectar padrões e regularidades, este é o nicho do *machine learning*. Cada padrão ajuda a entender o processo ou utilizar esses padrões para construir previsões, podendo assumir que no futuro próximo não seja tão diferente do passado com base nos dados coletados a previsão futura pode estar correta.

Aplicar um método de *machine learning* para uma grande base de dados é chamado de mineração de dados. A analogia é que um grande volume de terra e matéria prima é extraída para ser minerada que quando processado leva a uma pequena quantidade de um material muito valioso. A similaridade na mineração dos dados é que para um grande volume de dados é construído com modelo simples de uso valioso, tendo alta precisão preditiva.

Os módulos que estão em uso pela instância Moodle do nosso estudo em questão são dois tipos algoritmos: associação e clusterização.

Para o tipo de algoritmo de associação, foi utilizado o *Apriori* que é utilizado na mineração da frequência do conjunto de dados de forma totalmente supervisionada, ou seja, os dados não precisam ser rotulados, são executados de forma exaustiva por isso fornece resultados satisfatórios para explorar todas às regras dentro da confiança especificada. Neste estudo de caso foi aplicado o algoritmo de associação, de forma ao varrer os módulos que estão ativos na base, através de suas regras, possam nos apoiar no entendimento de comportamentos e

ajudar em estratégias futuras dentro do uso da plataforma. Abaixo, na figura 02 um exemplo simplificado de como o algoritmo *Apriori* trabalha:

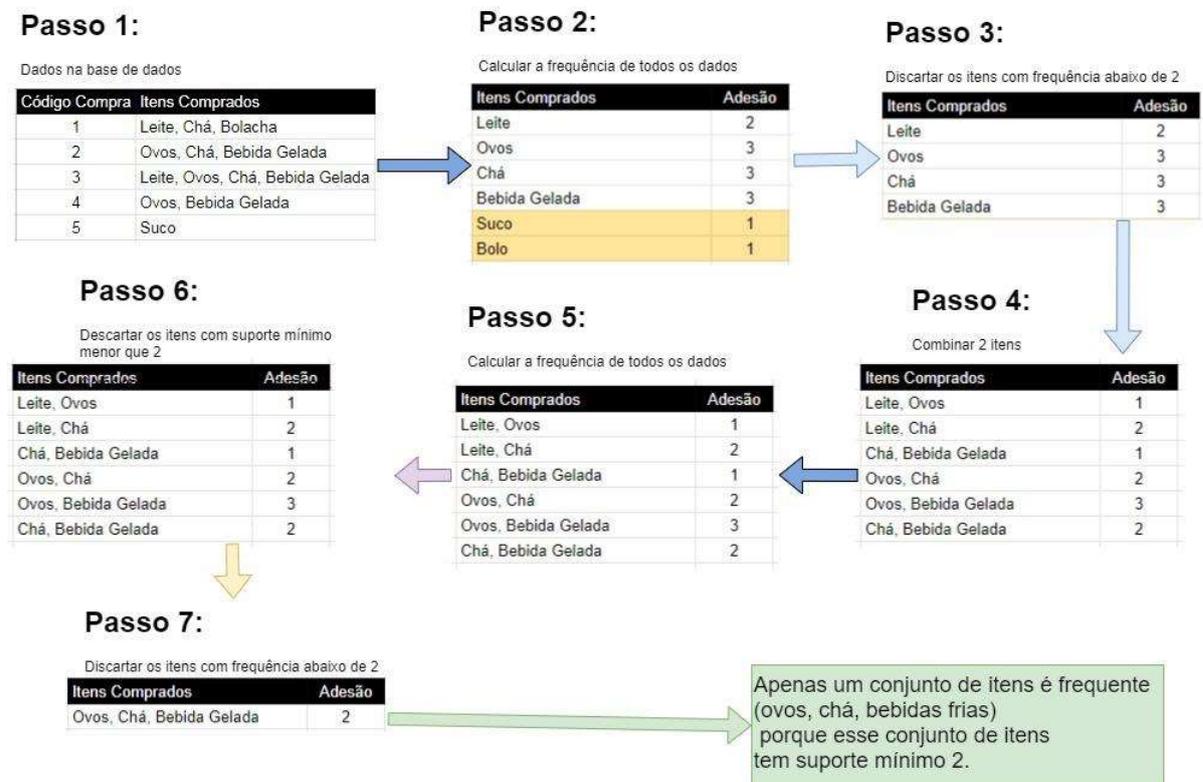


FIGURA 02: Como funciona o algoritmo *Apriori* (Traduzido)  
 fonte: <https://t4tutorials.com/apriori-algorithm-in-data-mining-with-examples/>,  
 acessado em 17/09/2019

Para o algoritmo de clusterização, foi utilizado o EM (*Expectation-Maximization*), explicado e publicado por Arthur Dempster, Nan Laird e Donald Rubin em 1977, foi utilizado para encontrar os parâmetros máximo de semelhança próxima nos casos onde a latência das variáveis envolvidas e os dados que estão faltando ou incompletos é utilizado para base de dados de *clusters* não supervisionados.

Passos realizados, conforme figura 03:

1. Dado os dados incompletos para considerar como dados para iniciar;
2. Passo Expectativa (E - step): usando os dados disponíveis observados do conjunto de dados, estime (adivinha) os valores dos dados ausentes;
3. Passo Maximização (M - step): os dados completos gerados após a etapa de expectativa (E) são usados para atualizar os parâmetros;
4. Repita as etapas 2 e 3 até a convergência.

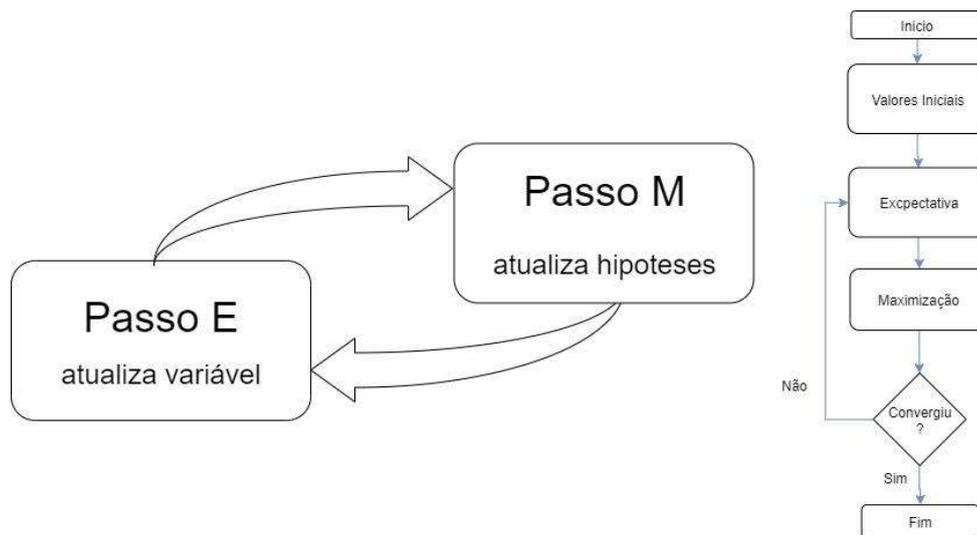


FIGURA 03: Como funciona o algoritmo EM (*expectation maximisation*) (Traduzido)  
 fonte: <https://www.geeksforgeeks.org/ml-expectation-maximization-algorithm/>,  
 acessado em 17/09/2019

## 2.4 Gestão do Projeto

Segundo o PMI (*Project Management Institute*) (2017), o projeto é um esforço temporário com tempo determinado, com escopo e recursos definidos. O projeto é único e não uma rotina operacional, mas configura às expectativas de pessoas em prol de um único objetivo.

A gestão do projeto é a aplicação dos conhecimentos, habilidades, ferramentas e técnicas para atividades do projeto para realizar os requisitos do projeto.

Às sessões incrementais para construção do produto são chamadas de *Sprints* e contém em linhas gerais um período de 30 dias por ciclo. Abaixo a figura 04 mostra um ciclo de entrega, chamado MVP (*Minimum Viable Product*) que é a entrega de um produto agregado para o usuário final por *Sprint*.

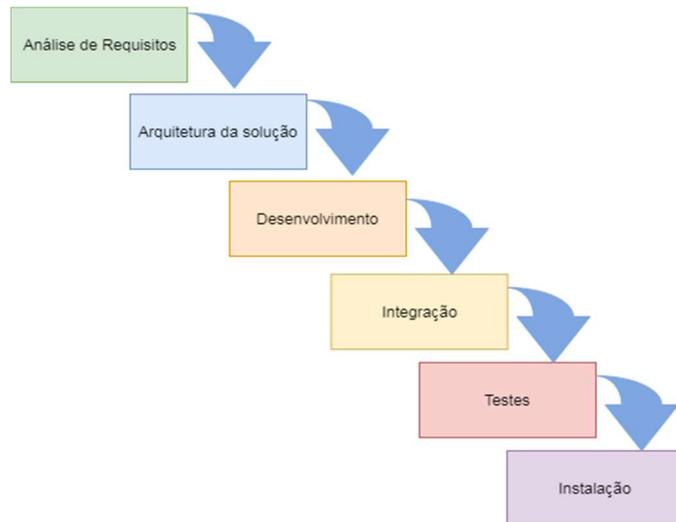


FIGURA 04: Ciclo de uma *Sprint*, repetido até o produto e evolução (traduzido).  
 fonte: JAMES (2017)

Geralmente a escolha do *Scrum* acontece em projetos de grande complexidade, onde o problema a ser resolvido é segregado em problemas menores de forma a gerar um valor de entrega para o cliente final, a figura 05 abaixo ilustra a abrangência da escolha de quando utilizar o *Scrum*:

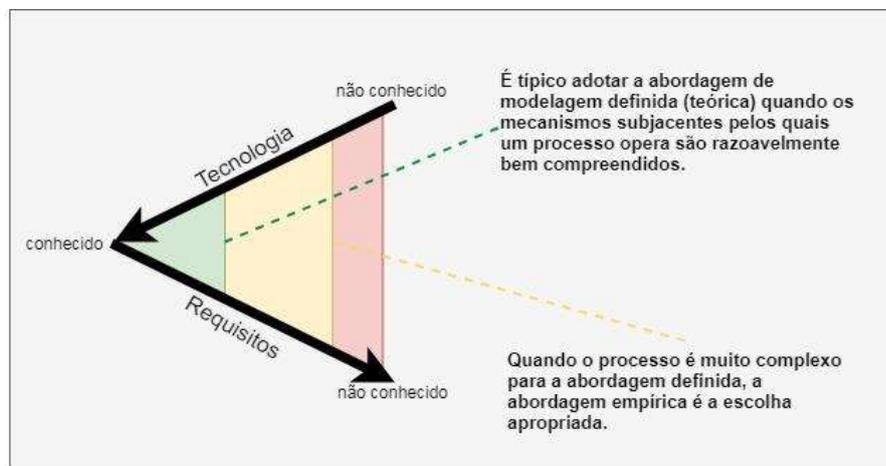


FIGURA 05: Escolha da metodologia *Scrum* em um projeto (traduzido)  
 fonte: JAMES (2017)

## 2.5 Ciclo de Vida do Software

Um gerenciamento do ciclo de vida da aplicação (ALM - *Application Lifecycle Management*), é o processo de transformar os requisitos em entregável tendo como resultado um software como produto. Os processos também incluem instalações, manutenções, suporte, evoluções, reformulações além de outros itens que

compõem também podemos considerar gestão de configuração e garantia da qualidade, aplicado em todo ciclo de vida do software.

O modelo utilizado foi modelo evolutivo (conforme figura 06), pois como trata-se de um estudo de caso, onde os requisitos não são todos conhecidos no início do desenvolvimento, onde a cada implementação de requisitos conhecidos e análise de resultados obtidos com o desenvolvimento relacionado aos objetivos principais, o software sofreu evoluções até ter todos os objetivos esperados, tendo assim a versão final.

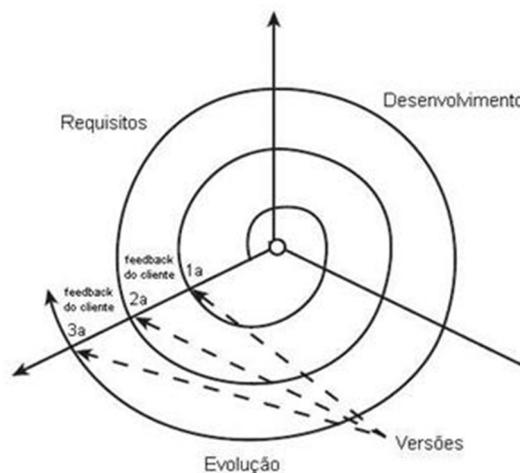


FIGURA 06: Modelo evolutivo usado no processo de engenharia de software  
fonte: SOMMERVILLE, 2003

## 2.6 Database

Para Jacsó (2019) termo banco de dados, refere-se a qualquer coleção de registros em formato eletrônico, onde eletrônico significa a capacidade de ler através do computador. Frequentemente os registros em uma base de dados representando alguma coisa, usualmente algo relacionado ao mundo real como por exemplo: um livro, uma publicação, uma peça de arte, um artefato arqueológico, ou mesmo uma pessoa (como registro de empregados, ou registro de pacientes nos casos de cuidado com a saúde). Em outros casos contudo, os registros são relacionados a itens como vários tipos de estatísticas, ou eles também compreendem a texto completo, de algo dos itens, tal como: relatórios técnicos, ou artigos de revistas e magazines.

A arquitetura do Moodle utiliza o banco de MySQL, Segundo Milani (2006), O MySQL é um banco de dados completo, robusto e extremamente rápido, com

todas as características existentes nos principais banco de dados pagos existentes no mercado. Uma de suas peculiaridades são suas licenças para o uso gratuito, tanto para fins acadêmicos como para realização de negócios, possibilitando que na maioria dos casos às empresas o utilizem livremente.

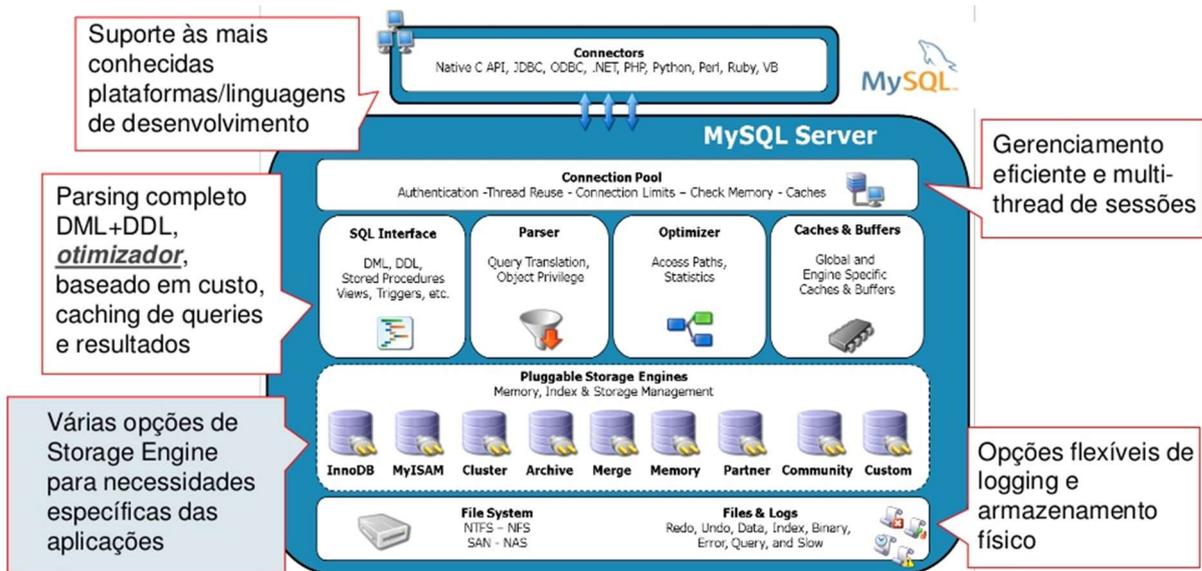


FIGURA 07: Arquitetura do banco de dados MySQL  
fonte: MySQL Brasil, 2013

## 2.7 ETL

Segundo Moss (2003), a proposta do processo de ETL é exibir as dependências entre todas as fontes de extração, executar utilitários de classificação e agrupamentos, transformação, tabelas e arquivos de trabalho temporário, tratamento de erros, atividades de reconciliação e a sequência de carregamento.

## 2.6 Dashboards

Segundo Eckerson (2011) organizações realmente precisam traduzir suas estratégias em: objetivos, métricas, iniciativas e tarefas customizadas para um grupo de pessoas e indivíduos dentro da organização. Isto promove informações oportunas e intuições que habilitam usuários do negócio para melhorar decisões, otimizar processos e planejar e trabalhar proativamente. Dashboard é realmente o sistema de gerenciamento de performance, comunicando objetivos estratégicos, e habilitando pessoas de negócios a mensurar, monitorar e gerenciar atividades e processos chaves necessárias para alcançar os seus objetivos.

No estudo de caso é utilizado *dashboards* para exibir o quanto está sendo utilizado cada módulo, nos módulos não utilizados, será informado às funcionalidades que o usuário está deixando de ter com a ferramenta, para concluir na sessão de mineração dos dados para cada módulo haverá um relatório disponível relacionado a cada tipo de algoritmo de mineração de dados, para ajudar na diagnóstico de uso da base.

Será utilizado o *framework* chart.js para exibição do gráfico relacionado ao uso da ferramenta, o *framework* utilizado é licenciado através de código aberto, baseado em javascript para uso em aplicações Web.

O modelo de gráfico que melhor representa o diagnóstico é o gráfico de barras horizontais que irá contribuir com uma leitura clara das informações disponibilizadas.

### **3 - O desenvolvimento do projeto**

Neste capítulo descreve-se o ferramental e ambiente adotado, explicando suas escolhas. Das ferramentas e *frameworks* utilizados, descreve-se o uso nas seguintes versões conforme Tabela 2:

Tabela 2: *Frameworks* utilizados no desenvolvimento

| Framework       | Versão  | Aplicabilidade  |
|-----------------|---|---|
| MySQL dump      | 10.13   | Versão do utilitário gerador do backup da base de dados.  |
| MySQL Distrib   | 5.7.23  | Versão da distribuição do Servidor MySQL, utilizado na base de dados.   |
| Linux           | ubuntu 16.04.1  | Versão do sistema operacional onde a instalação do Moodle objeto de estudo foi instanciada.   |
| POM             | 4.0.0   | <i>Project Object Model</i> ferramenta utilizada no projeto Maven Java, contendo os <i>archetypes</i> para os componentes Web e informações de configuração de estrutura e <i>frameworks</i> .  |
| JUNIT           | 5.4.2   | <i>Framework</i> responsável para fazer a gestão de testes unitários, cujo o objetivo é garantir qualidade funcional do código.   |
| Servlet         | 2.5   | Programa na linguagem Java capaz de trabalhar em um servidor de aplicação através do modelo de requisições e respostas de informações definidas através do protocolo HTTP ( <i>Hyper Text Transfer Protocol</i> , comunicação WEB entre cliente e servidor. ).  |
| Weka            | 3.8.2   | Coleção de algoritmos utilizados em aprendizado de máquina, para realização de atividades de mineração de dados.  |
| MySQL Connector | 8.0.16  | Conector Java utilizado na comunicação entre a aplicação Java e o banco de dados MySQL.   |
| Maven           | 2.6   | Permitir a compreensão completa do esforço do desenvolvimento, os principais objetivos são: <ul style="list-style-type: none"> <li>- facilitar o processo de construção;</li> <li>- construir um sistema uniforme;</li> <li>- utiliza às melhores práticas de desenvolvimento;</li> <li>- transparência na migração de novas versões de <i>frameworks</i>;</li> </ul> |
| Eclipse         | neon.3  | IDE ( <i>Integrated Development Environment</i> ), utilizado para construir efetivamente a aplicação.   |
| Tomcat/Apache   | 8.0/7.07.0  | Servidores de aplicação para uso local para desenvolvimento e remoto pelos usuários finais.   |
| GZip            |   | Utilitário para instalação da aplicação no servidor externo.  |
| JDK             | 8   | JDK ( <i>Java Developer Kit</i> ) ambiente para desenvolvimento de aplicações utilizando a linguagem de programação JAVA.   |
| VirtualBox      | Version 6.0.10<br>r132072<br>(Qt5.6.3)<br>Sistema Operacional:<br>Ubuntu 16.04<br>LTS | É uma plataforma de virtualização de aplicações. Estende a capacidade do seu computador existir em qualquer sistema operacional, dentro de uma máquina virtual, podendo ser mais um computador ao mesmo tempo.  |

Fonte: próprio autor

A realização das atividades para alcançar os objetivos deste estudo de caso, foi criado o *landscape* de *frameworks* e ferramentas com o intuito de trazer às melhores práticas no desenvolvimento e gestão da aplicação.

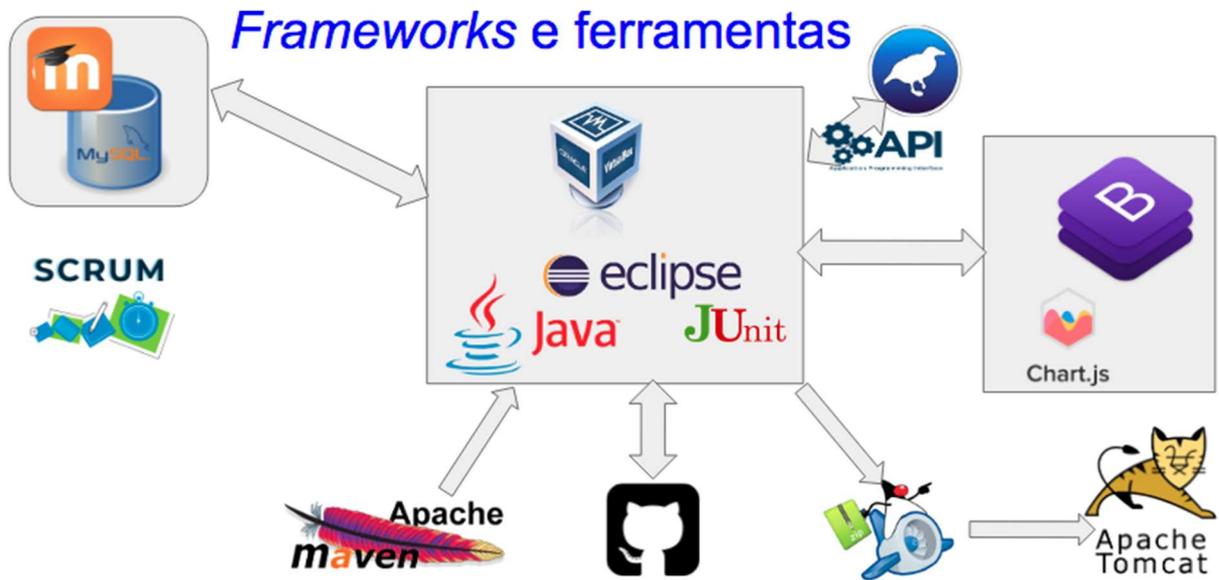


FIGURA 08: Arquitetura de ferramentas e *frameworks* utilizados na aplicação  
fonte: próprio autor

### 3.1 Gestão de Projetos

Devido a indefinição dos requisitos por se tratar de um estudo, os requisitos evoluíram ou ainda outros a serem implementados foram gerados em decorrência dos fatores de sucesso ou não, de forma a termos que tratar um ciclo de vida de software utilizado o módulo evolutivo, para orquestrar o dinamismo das atividades o SCRUM foi a metodologia escolhida para realizar a gestão do projeto, devido o fato de poder fracionar a entrega final, como se tratava de 3 três objetivos, fizemos o processo como cada objetivo distribuído em uma Sprint.

A figura 09 a seguir demonstra o requisito (objetivo 1) em andamento e os demais na fila aguardando a conclusão, após o término dos 3 objetivos implementados o objetivo central está concluído.

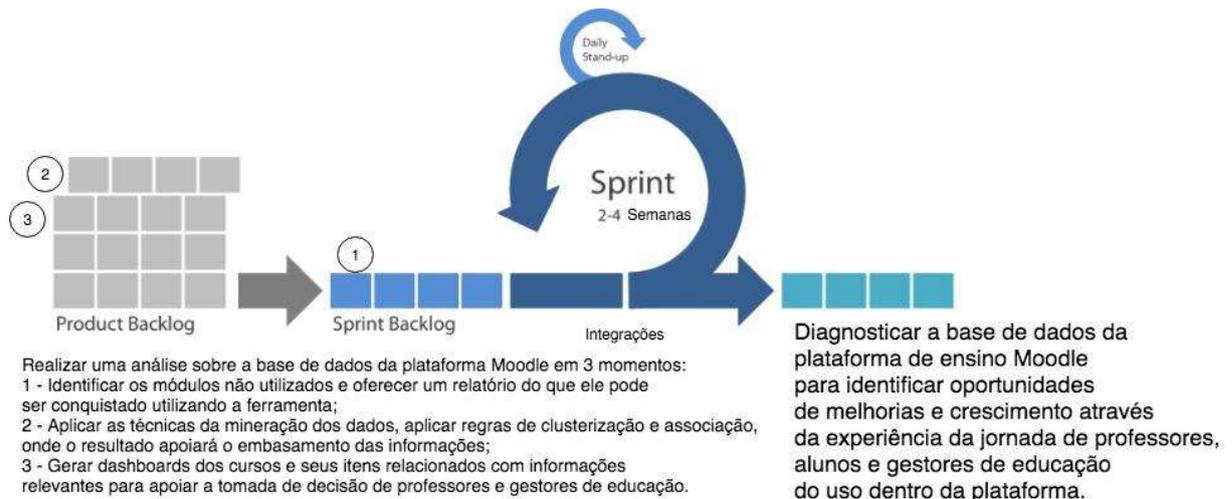


FIGURA 09: Metodologia SCRUM adotada no projeto (adaptado)

fonte: <https://www.zup.com.br/blog/scrum-conceitos-e-como-aplicar>, acessado em 17/09/2019.

Javadoc é a ferramenta que foi utilizada para gerar a documentação técnica relacionada a descrição de: classes, métodos, construtores, interfaces e campos.

A partir dos comentários feitos durante o desenvolvimento no código a ferramenta faz construção do documento no formato html, abaixo exemplo (figura 10) da estrutura da documentação técnica utilizada na elaboração da ferramenta desenvolvida.



FIGURA 10: Aplicação do Javadoc para geração da documentação técnica

fonte: próprio autor

## 3.2 ETL

Para realizar a extração, transformação e carregamento das informações referente a base de dados para disponibilizar os dashboards e relatórios, foi necessário estruturar o processo de ETL conforme figura 11:

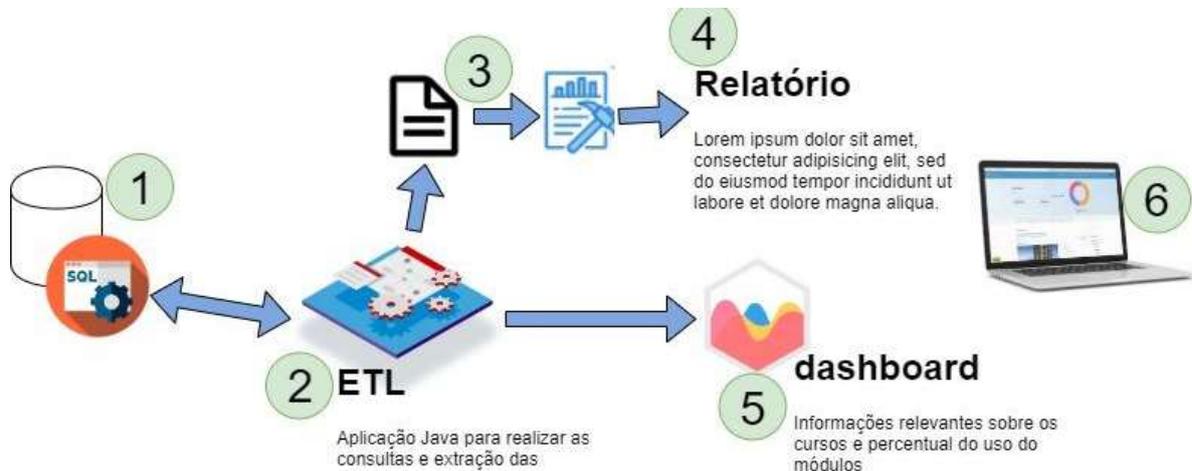


FIGURA 11: Estrutura do processo de ETL no projeto  
fonte: próprio autor

Uma explicação inicial sobre as etapas do processo:

1. Utilizado SQL para fazer a consulta às informações da base de dados, do estudo de caso;
2. Programas Java realizar criados para gerar as informações para o gráfico, gerar os arquivos do tipo arff para realização da mineração dos dados,
3. Programas Java executam a biblioteca do Weka para realizar a mineração dos dados e gerar os relatórios dos algoritmos de associação e clusterização;
4. Versão do relatório para consulta na Web disponibilizada;
5. Informações sobre os módulos em operação para geração de *dashboards* são providos;
6. Apresentação para o usuário final na Web dos relatórios de mineração de dados e informações de uso dos módulos.

### 3.3 Gerenciador de *frameworks* de desenvolvimento

A “ferramenta de construção” assim chamado por muitos usuários, é uma ferramenta que faz a implantação de artefatos no código fonte da aplicação. Como explicar a importância do MAVEN no projeto? O MAVEN é uma ferramenta de construção focada em reprocessar, compilar, empacotar, testar e distribuir o produto software. Além disso a função do MAVEN é abranger o *Project Object Model* (O POM é a unidade fundamental para se trabalhar com MAVEN. Em um arquivo no formato XML, que contém as informações de *frameworks* e versão com detalhes de configurações usados pelo MAVEN para construir o projeto) que configura os padrões, o ciclo de vida do projeto, sistema de gestão de dependências, a figura 12 demonstra a ordem de execução das tarefas do ciclo de vida usando MAVEN.

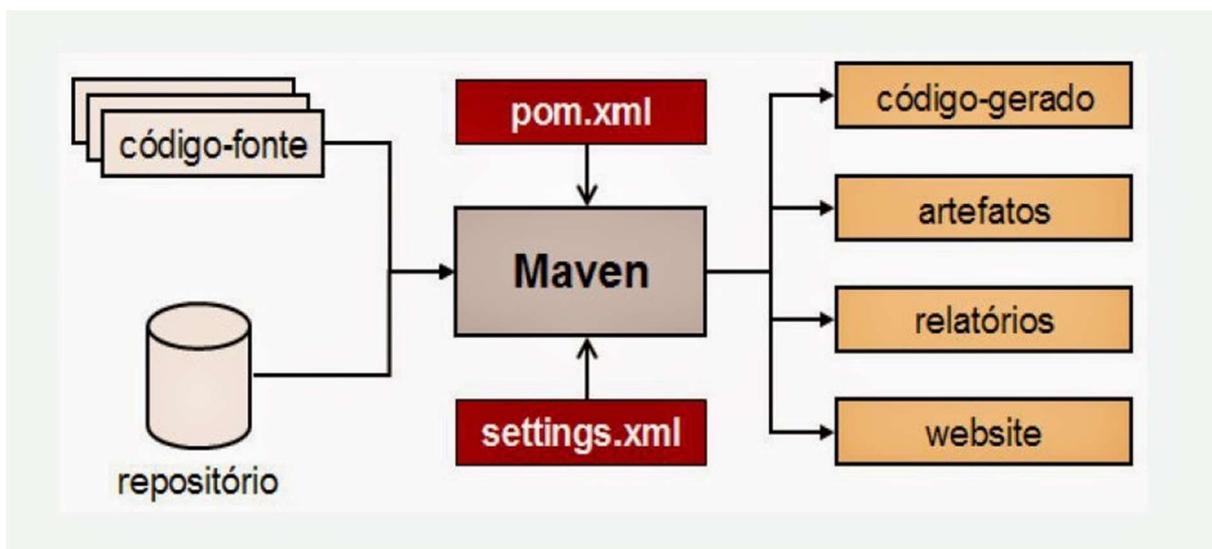


FIGURA 11: Ordem de execução das tarefas do ciclo de vida usando MAVEN  
fonte: <https://thallitaceleste.blogspot.com/2014/05/maven-repository.html>,  
acessado em 13/09/2019

Na aplicação que está desenvolvida objeto de nosso estudo utilizamos o MAVEN, no exemplo a seguir o arquivo POM está configurado para utilizar os frameworks e ferramentas: Junit na versão 5.4.2, Servlet na versão 2.5, Weka na versão 3.8.2, Conector do banco MySQL 8.0.16, versão do Maven 2.6.

### 3.4 Automatizando execução de testes unitários

Normalmente os programadores devem testar seus códigos. Testar utilizando o JUnit não é exatamente tão diferente do que os programadores já fazem hoje. É um caminho diferente de como fazer, uma forma automática de fazer, a diferença está em checar um resultado esperado e poder submeter uma bateria de testes, de forma a utilizar a ferramenta para monitorar se aplicação testada está se comportando com os resultados esperados conforme figura 12. Neste projeto utilizamos o JUnit na versão 5.4.2, onde foi escrita às classes de testes a fim de realizar testes básicos como por exemplo relacionados a conexão com banco de dados e *status* da conexão.

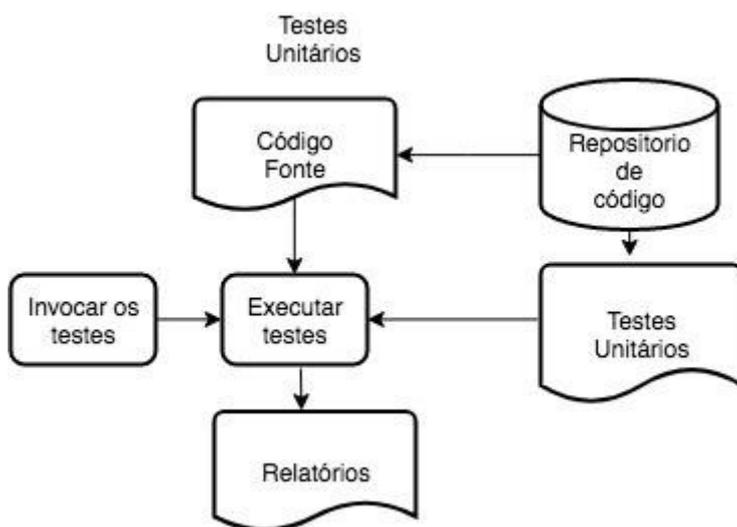


FIGURA 12: Arquitetura, aplicação e arquitetura utilizando JUnit (tradução) fonte:<https://www.testingdocs.com/junit-unit-testing-framework/>, acessado em 13/09/2019

Na aplicação desenvolvida utilizamos o JUnit na versão 5.4.2, conforme configurado no POM (Maven) do projeto, para automatizar os testes das classes criadas, com o objetivo de garantir o funcionamento adequado, dentro do esperado como resultado da aplicação.

### 3.5 Estrutura da solução técnica

A aplicação desenvolvida objeto de deste estudo é uma aplica Java Web. Foi utilizado o arquétipo Maven Webapp, para nortear o desenvolvimento da aplicação.

A figura 13 exibe a estrutura da aplicação de com relação a distribuição dos itens que compõem a aplicação utilizando arquétipo.

```
1. project
2. |-- pom.xml
3. |-- src
4.   |-- main
5.     |-- webapp
6.       |-- WEB-INF
7.         |-- web.xml
8.         |-- index.jsp
```

FIGURA 13: maven-archetype-webapp é um arquétipo qual gera um exemplo de um projeto webapp Maven

fonte: <https://maven.apache.org/archetypes/maven-archetype-webapp/>

O arquétipo é um *kit* de ferramentas para um *template* de um projeto Maven, é definido originalmente como padrão ou modelo do qual outras coisas do mesmo tipo são feitas. O uso de um arquétipo promove um grande caminho para que desenvolvedores facilmente em um caminho de melhores práticas empregadas em seus projeto, sua organização.

### 3.6 Configurando um ambiente desenvolvimento integrado

Um ambiente de desenvolvimento integrado (IDE - *Integrated Development Environment*), é uma aplicação que facilita o desenvolvimento. Em geral uma IDE é uma interface gráfica projetada para combinar as ferramentas necessárias às mãos. Às funções mais comuns são *debugar* o código, maximizar produtividade da construção e poder alterar entre outras aplicações.

Para o desenvolvimento do projeto foi utilizado o Eclipse como IDE, pois é possível customizar instalando apenas os serviços necessários.

No projeto foi instalado os gestores de testes unitários, arquétipos para projetos Web, gerenciamento de frameworks e o servidor de aplicação local, além claro a construção efetiva de codificação A figura 14 ilustra o uso de forma macro.

Pode-se facilmente combinar o suporte a linguagem com outras características dentro dos pacotes padrão, bem como agregar outros itens de forma a estender de forma ilimitada às customizações através do Eclipse *Marketplace*.

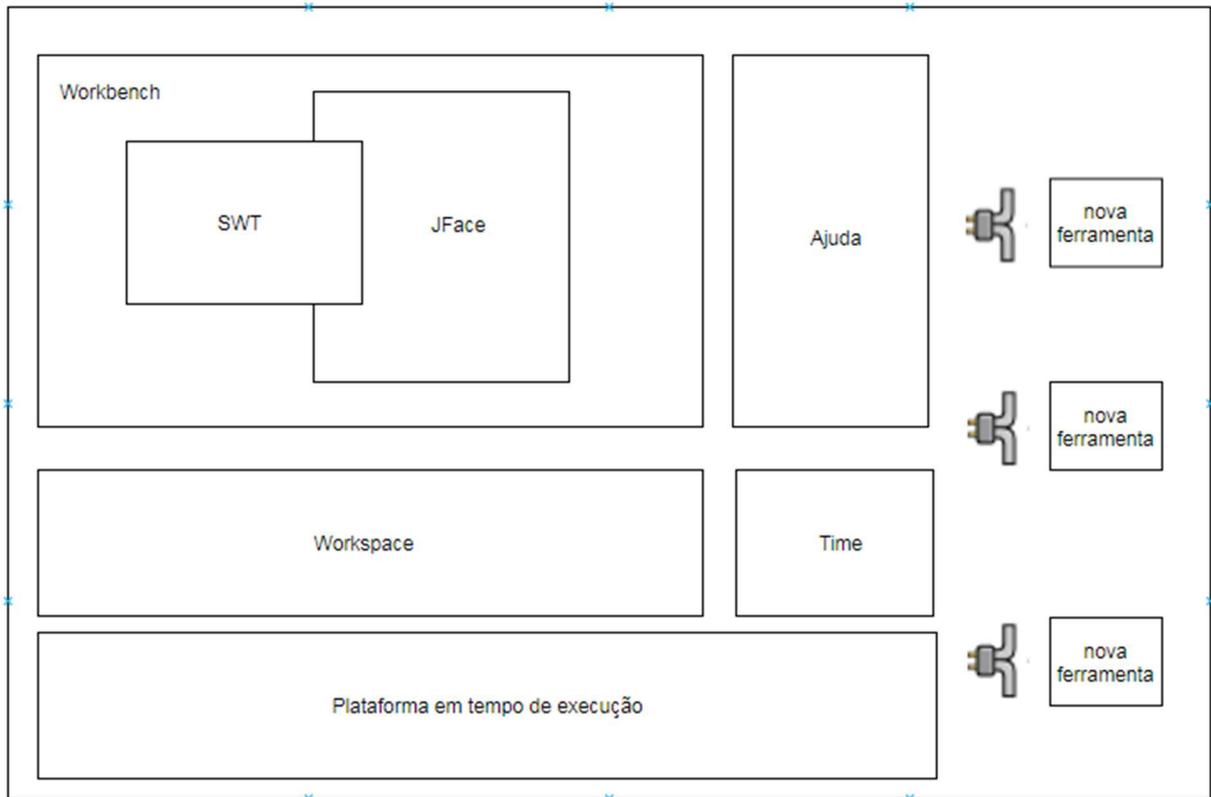


FIGURA 14: Arquitetura Macro da IDE Eclipse (traduzido)  
 fonte: <https://www.eclipse.org/forums/index.php/t/446510/>

### 3.7 Gerenciamento de código fontes

Github é uma plataforma de armazenamento de código para colaboração e versionamento do código, através do Github é possível trabalhar em time para construção de um projeto (conforme figura 15).

Em nosso projeto utilizamos o Github como repositório de versionamento de código, onde está disponível para comunidade de desenvolvedores e acadêmica para evolução futura, onde independente da estação cliente através dessa ferramenta associada a extensão Egit (instalada na IDE Eclipse na sessão *MarketPlace*) foi possível realizar os desenvolvimentos independente da estação trabalho, bem como ter todo o histórico de alteração e evolução do código.

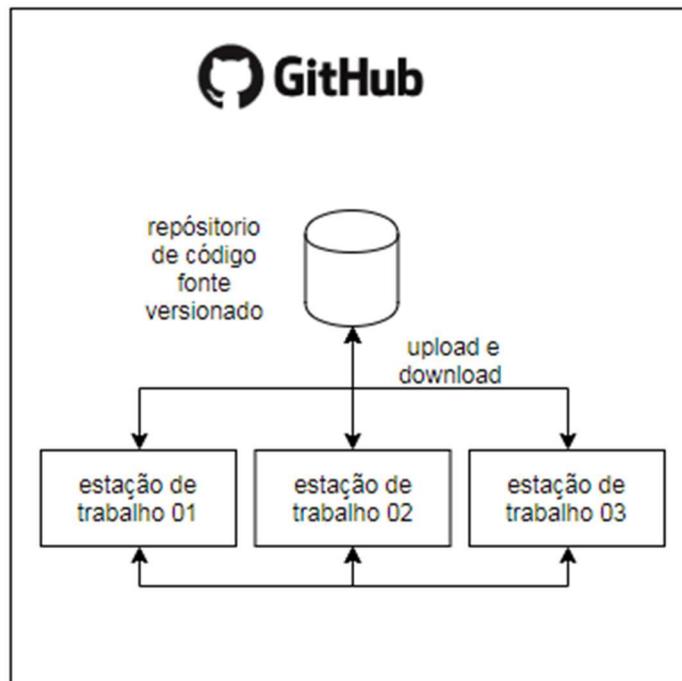


FIGURA 15: Macroprocesso demonstrando o uso do Github  
 fonte: <https://www.toolsqa.com/git/difference-between-git-clone-and-git-fork/>

### 3.8 Configurando o servidor de aplicação

O projeto servidor HTTP Apache é um esforço para o desenvolvimento e manter com licenciamento de código aberto para sistemas operacionais diversos, cujo principais objetivos são um servidor eficiente e extensível para prover serviços HTTP em sincronia com os padrões atuais de HTTP.

No projeto foi necessário configurar pastas de arquivos temporário com o devido permissionamento, pois arquivos gerados para realização e para download do usuário final, relacionados a mineração de dados estão armazenados em pasta local.

O servidor HTTP precisa suportar web services, aplicações Java J2EE, no estudo de caso a camada de apresentação e requisição e retorno de serviços (utilizando servlet) está disponível via internet armazenado em servidor HTTP, a figura 16 ilustra o funcionamento do servidor.

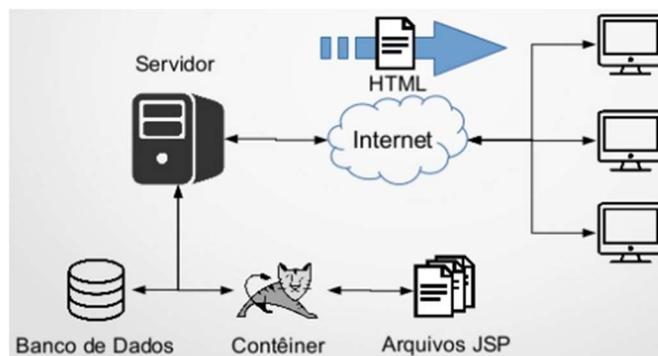


FIGURA 16: Visão geral do funcionamento da solicitação do usuário ao servidor de aplicação e seu retorno  
 fonte: <https://pt.slideshare.net/ayaherfc/slides-do-minicurso>, acessado em 18/09/2019.

### 3.9 Gerenciamento de armazenamento dos dados na plataforma Moodle

MySQL é o sistema de gerenciamento de banco de dados com licenciamento de código aberto mais popular, é desenvolvido, distribuído e suportado pela *Oracle Corporation*.

Uma base de dados é uma coleção estruturada de dados, isto pode ser desde uma simples lista de uma galeria de imagens ou um grande número de informações em uma rede corporativa. Para adicionar, acessar e processar dados armazenados, você precisa de um sistema de gerenciamento de dados. Os dados são organizados de forma relacional separada em algumas tabelas de forma a garantir a performance da consulta aos dados (figura 17).

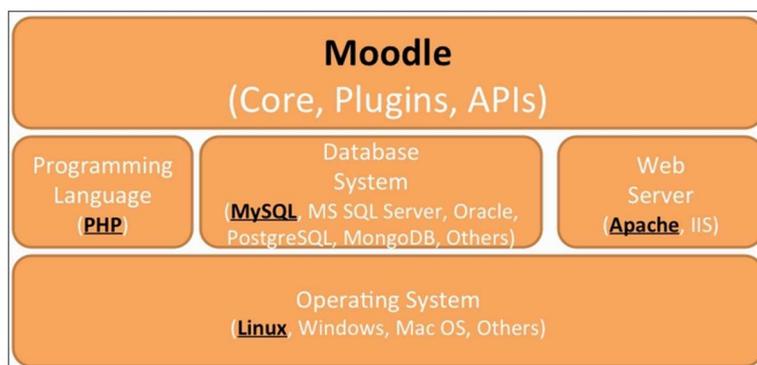


FIGURA 17: Arquitetura LAMP (Linux, Apache, MySQL e PHP) do Moodle  
 fonte: [https://subscription.packtpub.com/book/hardware\\_and\\_creative/9781783289714/2/ch02lv1sec15/moodle-architecture](https://subscription.packtpub.com/book/hardware_and_creative/9781783289714/2/ch02lv1sec15/moodle-architecture), acessado em 18/09/2019

O objeto do nosso estudo é a base de dados do Moodle, utilizando o servidor MySQL de banco dados.

### 3.10 Utilizando os recursos de mineração de dados

Em nosso estudo de caso é utilizado o Weka Java API, para incorporar a funcionalidade de mineração de dados na aplicação, executando sobre cada módulo em uso da instância Moodle objeto do estudo de caso os algoritmos *Apriori* (Associação) e o EM (clusterização).

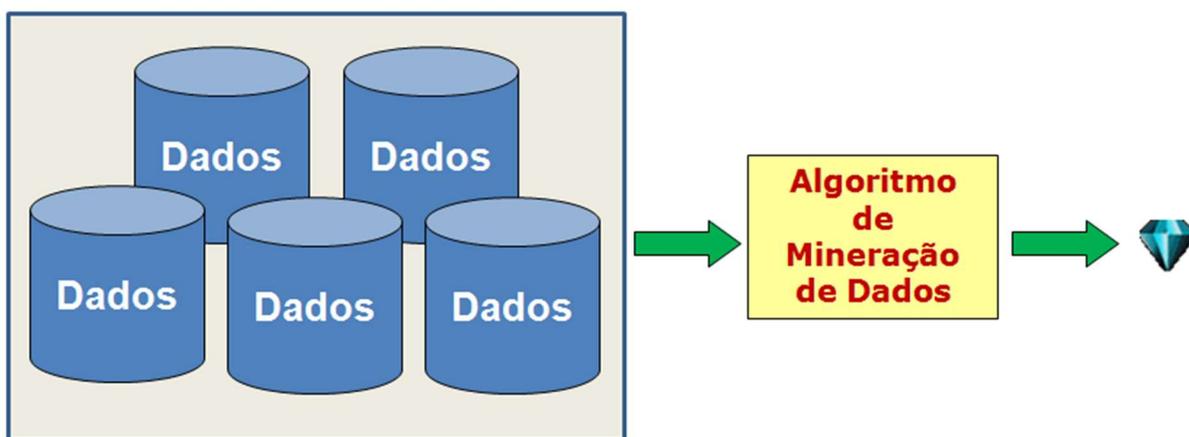


FIGURA 18: Idéia geral da mineração de dados

fonte: <https://www.devmedia.com.br/mineracao-de-dados-na-pratica-com-weka-api-revista-sql-magazine-107/26841>, acessado em 19/09/2019

### 3.11 Dashboards

Um *dashboard* de dados é uma ferramenta de gerenciamento de informação para exibir no formato de gráficos análises, métricas, pontos de monitoramento, saúde dos negócios.

O Chart.js é uma ferramenta com licenciamento MIT, e está sendo utilizado para geração dos gráficos dos *dashboards* em nossa ferramenta.

Para que o usuário tenha uma experiência amigável ao acessar a aplicação o *framework* Bootstrap para apoiar no *Look and Feel* da aplicação.

## 4 - Desenvolvimento

Neste capítulo é descrito as etapas de trabalho, o desenvolvimento em si, método.

### 4.1 Metodologia de desenvolvimento

A metodologia utilizada para realizar a gestão do projeto é o *Scrum*, que permite o desenvolvimento incremental do projeto considerando uma ou mais funcionalidades conduzidas por times auto organizados.

Este tipo de metodologia promove uma estrutura de papéis, reuniões, regras e artefatos. Os times são responsáveis por criar e adaptar o processo dentro do *framework*.

### 4.2 Interação das fases

Através do modelo evolutivo à medida que os requisitos evoluem ou ainda ficam mais claros, a definição do requisito é concluída dando andamento às atividades de desenvolvimento, no caso quando um problema surgia os requisitos eram novamente revistos e evoluídos, podendo fazer a revisão do desenvolvimento até o resultado produto.

Para realização do desenvolvimento da aplicação que é o produto deste estudo de caso foi realizado a configuração inicial do ambiente que consiste em:

- criação da máquina virtual onde foi criado o ambiente de desenvolvimento;
- criação do ambiente de desenvolvimento;
- instalação e configuração das ferramentas e *frameworks* utilizados;
- contratação de serviço de hospedagem externa para acesso via internet do sistema desenvolvido;
- restauração da base de dados objeto deste estudo de caso, deixando acessível os dados no ambiente de desenvolvimento.

Após a configuração inicial do sistema, foi realizado o próximo passo foi análise da base de dados e busca de documentação relacionada de que forma o banco de dados está estruturado para então implementar a codificação da solução.

A tabela 03 exibe o número de tabelas por módulo, que foram divididas conforme a documentação do produto em módulos para a realização da atividade.

TABELA 03: Número de Tabelas por módulo

| Módulo                  | Tabelas |
|-------------------------|---------|
| Tipos de Questões       | 22      |
| Competencias            | 18      |
| Mensagens               | 15      |
| Workshop                | 14      |
| Atribuição              | 13      |
| Classificar             | 11      |
| Testes                  | 11      |
| Conteúdo Compartilhável | 11      |
| Usuarios e Perfis       | 11      |
| Classificação Avançada  | 10      |
| Curso                   | 10      |
| Perfis                  | 10      |
| Análise                 | 9       |
| Forum                   | 9       |
| Reconhecimento          | 8       |
| Lição                   | 8       |
| Banco de Questões       | 8       |
| Wiki                    | 7       |
| Pesquisa                | 4       |

fonte: próprio autor

Com base nestas informações iniciais, divide-se o diagrama completo de relacionamento do banco de dados, em diagramas menores, em 21 módulos iniciais, cujo objetivo foi coletar as tabelas e seus relacionamentos, para iniciar o processo de consulta das informações por módulo, a seguir a figura 19 exibe o exemplo de agrupamento do módulo *analytics*.

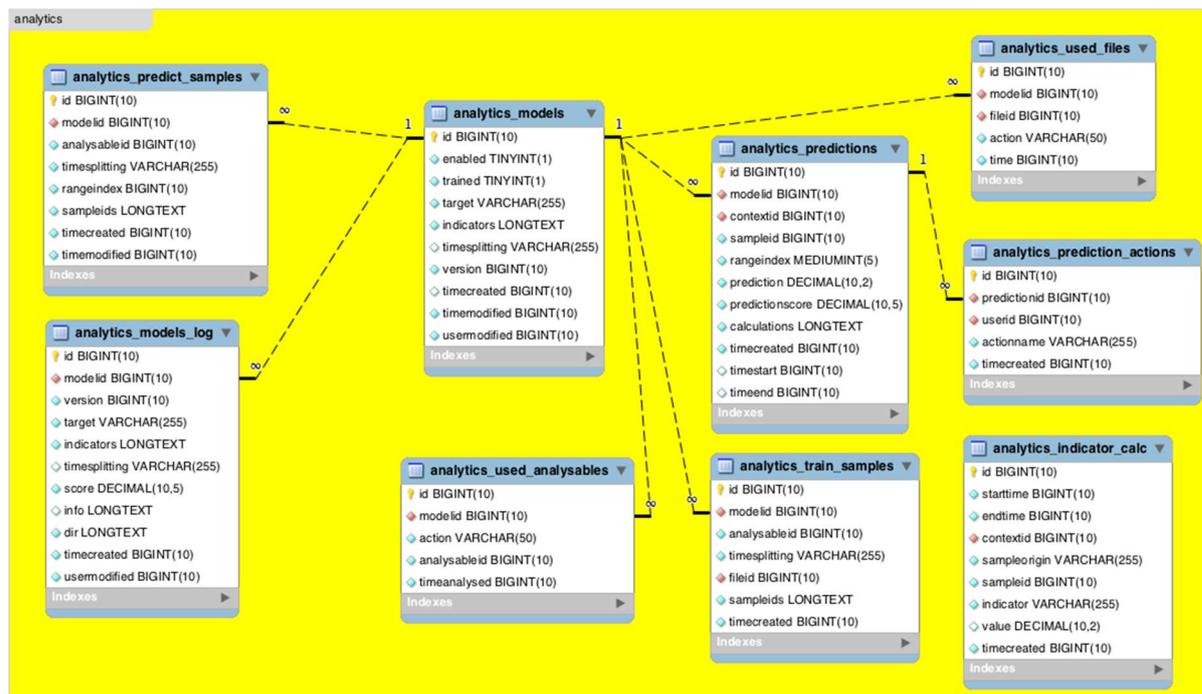


FIGURA 19: Módulo *Analytics*

fonte: <http://www.examulador.com/er/components/analytics.html> 20/09/2019

Foi realizado esse mesmo trabalho sobre mais 20 módulos, onde após concluído a identificação dos módulos foi feita a identificação das tabelas que compõem os módulos, para que pudesse ser possível quantificar o quanto o módulo está sendo usado, caso o cálculo fosse igual a 0%, significaria que o módulo está sem uso, neste caso com base na documentação da ferramenta foi criada uma mensagem sobre o que o módulo faz para fomentar o uso do módulo na instância.

### 4.3 Processo de contabilização dos módulos

Para realizar a contabilização dos módulos, foi necessário às seguintes etapas:

1. Identificamos com base na documentação de banco de dados do Moodle quais são os módulos e às tabelas que fazem parte de cada módulo;
2. Realizada uma instrução de consulta aos dados da tabela, com o objetivo de identificar se existem dados na tabela;
3. Realizada uma contabilização do número de tabelas total;
4. Aplicada a fórmula de uso = (tabelas com dados/todas tabelas do módulo) / 100;

5. Informação registrada no dashboard, para módulo com uso igual a 0%, não registrada no dashboard e sim no relatório de recomendação de uso.

#### **4.4 Processo de processamento de mineração de dados**

As informações para realização das atividades dos algoritmos de mineração de dados, para que pudesse processar, neste caso, foi identificado manualmente, com base na documentação do Moodle, em cada módulo, qual é a principal tabela, após esse passo foi gerado um programa que faz a leitura da tabela no banco de dados e cria o arquivo com extensão .arff, e em seguida é utilizado este arquivo .arff para gerar o relatório do resultado na mineração de dados no formato .txt, para este relatório é foi feita utilizada os algoritmos *Apriori* de associação e EM de clusterização.

#### **4.5 Apresentação dos resultados para o usuário final**

Com base nas informações dos módulos, o próximo passo foi gerar o gráfico com informações de uso dos módulos e para os módulos que não são utilizados, uma recomendação para motivar o uso. Para a mineração dos dados foi criado uma ferramenta para download do .txt para cada tipo de algoritmo (associação e clusterização) para realizar consulta do relatório de mineração dos dados.

## 5 - Resultado

Com o *dashboard* implementado, foi possível disponibilizar uma visão geral de como a ferramenta está sendo utilizada, bem como promove informações relacionada aos módulos não utilizados, para fomentar o uso do módulo, abaixo a informações promovidas do objeto deste estudo de caso.

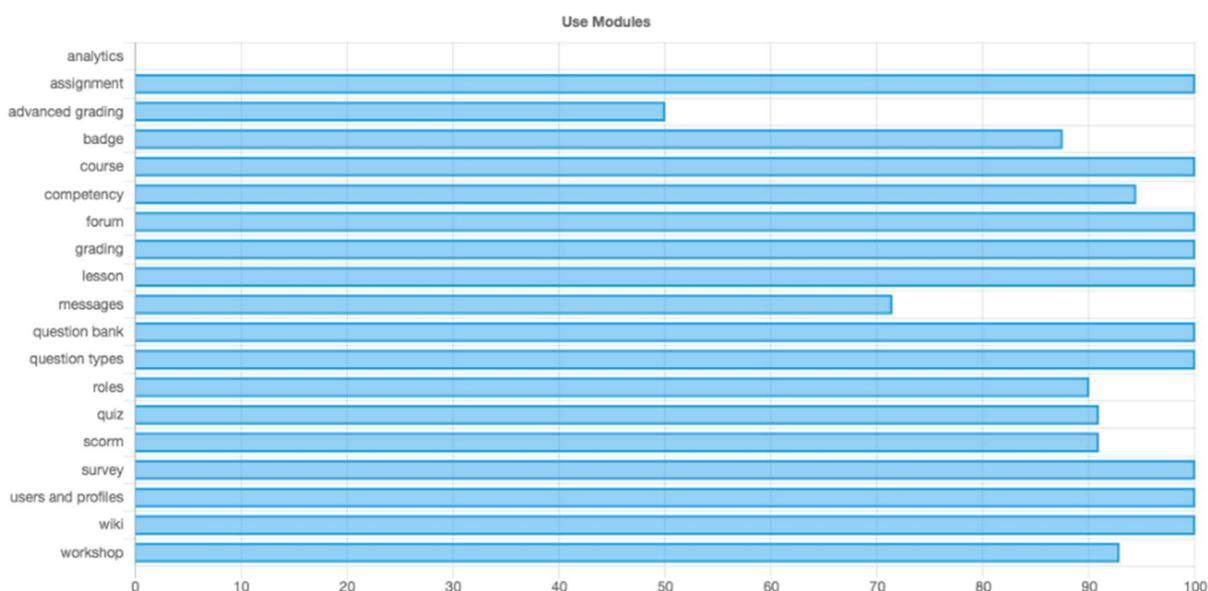


FIGURA 20: Exibição da contabilização dos módulos na ferramenta  
fonte: próprio autor

Exemplo do relatório para fomentar o uso da ferramenta:

More information about **unused** modules

The module **analytics** is not use.

When this module **is not use**, you left of the use the follow functions: **Often in the past, learning analytics systems have attempted to analyze past activities to predict future activities in real time. With Moodle Learning Analytics, we are more ambitious. We believe a full learning analytics solution will help us not only predict events, but change them to be more positive.**

FIGURA 21: Exibição da contabilização de informações de módulos não utilizados  
fonte: próprio autor

Para realização do relatório de mineração o usuário pode selecionar qual módulo e solicitar o processamento. Após a conclusão os relatórios dos algoritmos de associação e clusterização são disponibilizados para download.

Select the module, after this click Load Reports button to generate two data mining reports: association (Apriori) and clustering (EM).

- analytics
- assignment
- advanced grading
- badge
- course
- competency
- forum
- grading
- lesson
- user & profiles
- question
- roles
- quiz
- scorm
- survey
- wiki
- workshop



FIGURA 22: Visão geral das etapas para processamento dos relatórios de mineração.

fonte: próprio autor

Com base nas informações do uso e os relatórios de mineração a tomada de decisão relacionada ao comportamento e o uso estão a disposição no formato web para os usuários.

O algoritmo de associação com sua grande gama de situações de aplicabilidade a respeito de características de consumo com o objetivo de maximizar relações de comportamento conforme os registros armazenados na base. Neste trabalho a aplicabilidade sobre cada módulo é apoiar o usuário da aplicação a ter informações que possa melhorar a experiência e aderência dos usuários o uso da ferramenta. Exemplo com o módulo de cursos, exibindo as 10 melhores regras encontradas:

1. newsitems=5 618 ==> summaryformat=1 618 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
2. summaryformat=1 618 ==> newsitems=5 618 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
3. showgrades=1 newsitems=5 617 ==> summaryformat=1 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
4. summaryformat=1 showgrades=1 617 ==> newsitems=5 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
5. newsitems=5 maxbytes=0 617 ==> summaryformat=1 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
6. summaryformat=1 maxbytes=0 617 ==> newsitems=5 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
7. newsitems=5 showreports=0 617 ==> summaryformat=1 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
8. summaryformat=1 showreports=0 617 ==> newsitems=5 617 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
9. showgrades=1 newsitems=5 maxbytes=0 616 ==> summaryformat=1 616 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)
10. summaryformat=1 showgrades=1 maxbytes=0 616 ==> newsitems=5 616 <conf:(1)> lift:(1) lev:(0) [0] conv:(1)

O algoritmo de clusterização na aplicação tem objetivo de definir a representação de padrões, feito uma seleção de subconjuntos com efetividade nos padrões disponíveis para descrever o padrão. Exemplo foi gerado 2 clusters e exibidos o resultado do campo *show reports*:

Number of clusters selected by cross validation: 2

Number of iterations performed: 0

```

                Cluster
Attribute      0  1
              (0.54)(0.46)
=====
showreports
0              337 283
1              2   1
[total]       339 284
```

A solução cumpriu os requisitos que foi proposto como objetivos como exibir informações relevantes sobre o uso dos módulos, o relatório de recomendação sobre os módulos que não estão em uso e para fechamento disponibilizado por módulo a possibilidade de ter relatórios de clusterização e associação para uma análise mais aprofundada.

## 6 - Conclusão

A contribuição deste estudo de caso para comunidade de usuários ao redor do mundo que utilizam a plataforma Moodle, tem a finalidade de disponibilizar uma ferramenta que pode apoiar a tomada de decisões com o mínimo de intervenção humana, através dos algoritmos de aprendizado de máquina que contribui com um modelo analítico, com neste estudo de caso é utilizado dois tipos de algoritmos o de associação e clusterização sobre o mesmo módulo agregado visões diferentes sobre o assunto.

O gráfico e relatório de uso disponibilizado tem a finalidade de contextualizar o usuário da plataforma rapidamente sobre os recursos utilizados e o quanto estão sendo utilizados utilizando uma escala de 0 a 100% com base na quantidade de tabelas utilizadas, não a quantidade de informações. Com base no documento do fabricante, para os módulos não utilizados é possível enfatizar oportunidades de melhoria.

Por fim o Moodle *analytics* é uma ferramenta que pode ser utilizada em qualquer instância Moodle e também pode ter evoluções pois todos seus componentes estão disponíveis no repositório de código para fazer customizações e evoluções personalizadas, os resultados neste estudo de caso é satisfatório nos aspectos de cumprimento ao objetivo proposto de poder fazer uma análise do uso dos módulos, com informações disponibilizadas através de dashboards e relatórios do tipo clusterização e associação de mineração de dados para uma análise mais aprofundada sobre o uso. A ferramenta também seguiu boas práticas de desenvolvimento e os testes com a base de dados objeto de nosso estudo obtém-se o resultado esperado.

## Referências bibliográficas

Bourque, P.; Fairley, R.E.; SWEBOK v3.0 - Guide to the Software Engineering Body of Business - John Wiley & Sons - 2011

Guia PMBOK - Sexta edição - Project Management Institute, 2017

Knowledge - IEEE computer society - 2019

Java Archive Downloads - Java SE 10. Disponível em:

<<https://www.oracle.com/java/technologies/java-archive-javase10-downloads.html>>.

Acesso em: 20 set. 2019.

Chicago and London - 2019

Moodle ER Diagram. Disponível em:

<<http://www.examulator.com/er/components/index.html>>. Acesso em: 20 set. 2019.

What Is a Servlet? - The Java EE 5 Tutorial. Disponível em:

<<https://docs.oracle.com/javaee/5/tutorial/doc/bnafe.html>>. Acesso em: 20 set. 2019.

Welcome to Oracle VM VirtualBox. Disponível em:

<<https://www.virtualbox.org/manual/ch01.html>>. Acesso em: 20 set. 2019.

What is Application Lifecycle Management - Learn the Tools & Methods. Disponível em:

<<https://www.inflectra.com/spirateam/highlights/understanding-alm-tools.aspx>>. Acesso em:

20 set. 2019.

What is Project Management? Disponível em:

<<https://www.pmi.org/about/learn-about-pmi/what-is-project-management>>. Acesso em:

20 set. 2019.

Escolha da metodologia *Scrum* em um projeto (traduzido). Disponível em:

[https://www.collab.net/sites/default/files/uploads/CollabNet\\_scrumreferencecard.pdf](https://www.collab.net/sites/default/files/uploads/CollabNet_scrumreferencecard.pdf).

Acesso em 20 set. 2019