

Mineração de dados educacionais para prever o desempenho escolar dos estudantes de física ao longo do Ensino Médio.

Ricardo Dias Pacheco Martins¹, Carlos Diego Nascimento Damasceno², Seiji Isotani³

Resumo

Com a crescente disseminação do uso de dados, cria-se uma janela de exploração em relação aos dados coletados e não trabalhados para adquirir conhecimento. Através de métodos como o KDD (Knowledge Discovery from Data) e EDM (Educational Data Mining) é possível realizar estudos que auxiliam no processo de ensino e aprendizagem, revelando informações relevantes e com potencial de responder questionamentos de como melhorar o desempenho escolar de um estudante. Este trabalho conduz de forma teórica e prática a aplicação das técnicas de mineração de dados no contexto da disciplina de física do Ensino Médio durante os três anos de uma mesma turma. O objetivo é prever de forma exemplificada por meio das árvores de decisão geradas pelo algoritmo J48 no WEKA, quais discentes têm chance de ficar de recuperação ou até mesmo reprovar.

Palavras-Chave: Mineração de Dados Educacionais, KDD, desempenho escolar

Abstract

With the increasing dissemination of data use, an exploration window is created in relation to the collected data and not used to acquire knowledge. Through methods such as KDD (Knowledge Discovery from Data) and EDM (Educational Data Mining) it is possible to carry out studies that assist in the teaching and learning process, revealing relevant information and with the potential to answer questions about how to improve a student's school performance. This work leads in a theoretical and practical way the application of data mining techniques in the context of high school physics discipline during the three years of the same class. The goal is to predict in an exemplified way with the use of decision trees that are created through the J48 algorithm in WEKA, which students have a chance to stay in recovery or even fail.

Keywords: Educational Data Mining, KDD, school performance

1 Pós-Graduando(a) em Computação Aplicada à Educação, ICMC-USP, <ricardomartins@usp.br>.
2 Co Orientador, Universidade de São Paulo, <damascenodiego@usp.br>.
3 Orientador, Instituto De Ciências Matemáticas e de Computação - USP, <sisotani@icmc.usp.br>.

1. Introdução

A possibilidade de contribuição pedagógica é fundamental devido ao crescente uso da tecnologia e a criatividade dentro do campo da educação. Nessa perspectiva, o trabalho tem por objetivo um entendimento melhor de como utilizar os dados educacionais gerados pelas instituições, assim criando indicadores de tendência, discussões e novos espaços a serem explorados segundo Vosgerau et al. (2014).

Como diz Witten (2005), vivemos em um mundo onde grandes quantidades de dados são coletados diariamente e analisá-los faz parte de uma necessidade importante. De acordo com Witten (2005), "estamos vivendo na era da informação", porém o ideal seria dizer que estamos vivendo na "era dos dados". Ainda de acordo com o autor, todos os dias, dados são gerados e despejados em nossas redes, armazenados em diferentes áreas do conhecimento e descartados sem ao menos levantar questionamentos de como podem influenciar em tomadas de decisões futuras.

Os dados atualmente estão sendo deixados de lado no campo educacional, o artigo pretende auxiliar os professores a decupar dados brutos através do conhecimento adquirido no campo da descoberta do conhecimento a partir de dados e aplicar o método do KDD (*Knowledge Discovery Data*) juntamente com o EDM (*Educational Data Mining*) para prever de forma simples, visual e prática como melhorar o desempenho escolar dos estudantes de Física do Ensino Médio ao longo do 1º, 2º e 3º anos. Para isso, foram coletados os dados de três professores e depois trabalhados para responder o questionamento de como identificar possíveis recuperações ao longo das primeiras unidades e semestre. A outra indagação é a respeito de como prever caso o estudante esteja com chances de reprovação. Romero et al. (2005) afirmam que a estratégia de predição pode e deve ser usada em pesquisas relacionadas a um conjunto de estudantes, assim beneficiando o campo educacional.

A partir deste contexto, o presente artigo tem como objetivo promover uma reflexão acerca da potencialidade do uso da mineração de dados no campo da educação dentro do Ensino Médio, apontando para um caminho conclusivo que mostre como o professor pode trabalhar esse desafio.

2. Fundamentação teórica

Baker, Isotani e Carvalho (2011) explicam que a uma das áreas de pesquisas em crescimento é a Mineração de Dados Educacionais (MDE), também conhecida como EDM, cujo o campo de pesquisa é realizar o processo de descoberta do conhecimento a partir da exploração dos dados produzidos e coletados a fim de realizar a decupagem e posteriormente produzir informações que sejam relevantes no ambiente educacional. A mineração de dados educacionais é uma ciência experimental e para ficar mais fácil o entendimento do termo, podemos fazer uma analogia com as pepitas preciosas encontradas ao meio de tanta matéria-prima. No campo educacional também podemos ou não encontrar muitas pepitas se o garimpo for bem feito e de forma organizada, pois o processo é essencial para obter conhecimento segundo Witten (2005). Da mesma forma que existe o garimpeiro, existe o cientista de dados, como também o professor

eufórico aspirante por busca de significados em meio a tanta informação gerada. Entretanto, para se aventurar nesse campo é necessário seguir algumas diretrizes para obter minimamente algum resultado. Witten (2005) explica que o processo de mineração de dados está aliado ao método KDD e que são praticamente sinônimos, enquanto Han et al. (2006) veem a mineração como parte principal dentro do processo de descoberta de conhecimento.

A seguir, podemos observar algumas técnicas de aplicação categorizadas na mineração de dados educacionais pela taxonomia proposta por Baker, Isotani e Carvalho (2011).

- Predição (*Prediction*)
 - Classificação (*Classification*)
 - Regressão (*Regression*)
 - Estimativa de Densidade (*Density Estimation*)
- Agrupamento (*Clustering*)
- Mineração de Relações (*Relationship Mining*)
 - Mineração de Regras de associação (*Association Rule Mining*)
 - Mineração de Correlações (*Correlation Mining*)
 - Mineração de Padrões Sequenciais (*Sequential Pattern Mining*)
 - Mineração de Causas (*Causal Mining*)
- Destilação de dados para facilitar decisões humanas (*Distillation of Data for Human Judgment*)
- Descobertas com modelos (*Discovery with Models*)

Baker, Isotani e Carvalho (2011) compartilham um pensamento que auxilia a entender a questão. Segundo os autores a área de predição tem o objetivo de desenvolver modelos que através de variáveis preditoras respondam a algo específico relacionado aos dados coletados e trabalhados. Uma das técnicas mais utilizadas é a classificação, onde a escolha da variável e a categorização dos indicadores são de suma importância para geração de uma representação visual conhecida como árvore de decisão.

Fayyad et al. (1996) apresentam o processo de KDD como um método eficaz para mineração e descobertas em diversas áreas do conhecimento. Primeiramente é necessário o levantamento dos dados, observação e compreensão para selecionar dados relevantes com o foco no objetivo proposto.

Os próximos passos de acordo com os autores citados anteriormente consistem com a limpeza e pré-processamento dos dados que engloba operações de limpeza para redesenhar as informações. Através desse passo, decidi-se estratégias para superar o tratamento de dados faltantes ou até mesmo mudanças de sequência caso necessário.

A etapa de transformação de dados é considerada fundamental para organizar as variáveis e ressaltar os atributos considerados úteis em relação ao objetivo, ou seja, por meio da transformação é possível categorizar, reduzir, segmentar e condensar os dados de forma efetiva para alcançar os resultados projetados e esperados com base no algoritmo de mineração escolhido.

A mineração de dados é o momento de exploração e seleção de modelo, escolhendo o algoritmo a ser utilizado para interpretar os dados anteriores já preparados. Por fim e não menos importante está a interpretação e avaliação dos modelos gerados, resultando em conhecimento a partir dos dados descobertos. A seguir, a figura 2.1. reproduz os passos sugeridos para aplicação.



Figura 2.1. Panorama dos passos que compõem o Processo de Descoberta de Conhecimento em Dados (KDD)

Fonte: Adaptado de FAYYAD et al. (1996).

Para a execução dos modelos, estudos e exploração necessária utiliza-se do programa Weka, apresentado por Witten (2005) em seu livro *Data Mining*. Weka (*Waikato Environment for Knowledge Analysis*) deu início em 1993, na Universidade de Waikato, Nova Zelândia e é um software Open Source que roda em Java e que tem por objetivo juntar diversos algoritmos a fim de facilitar a mineração de dados para fornecer interpretações dos dados coletados e trabalhados durante o processo do KDD. Witten (2005) também nos lembra que não existe um método 100% apropriado para encontrar todas as respostas durante a mineração de dados.

Vale ressaltar que o uso do programa WEKA necessita de uma certa preparação dos dados e também de um arquivo específico tipo ".ARFF" para trabalhar a técnica de classificação e depois a geração de uma árvore de decisão. Para o tratamento dos dados e descoberta a respeito do conhecimento escolar dos estudantes da disciplina de física do Ensino Médio será utilizado o algoritmo J48 que será melhor discutido nas próximas seções.

3. Trabalhos relacionados

As pesquisas realizadas por Alves et al. (2018) e Rabelo et al. (2017) possuem características que afiguram-se com o objetivo levantado no respectivo trabalho, pois buscam através da metodologia do KDD (*Knowledge Discovery Data*) criar um modelo para prever o desempenho de matemática e suas tecnologias no Enem e o desempenho de alunos de EaD em Ambientes Virtuais de Aprendizagem utilizando a técnica de mineração de dados. Os dois artigos citados utilizam também o software WEKA para colaborar com o processo de mineração de dados através do algoritmo J48, resultando em uma árvore de decisão com base na categorização das variáveis. No trabalho de Alves et al. (2018), observa-se que a base de dados referente a Matemática e suas Tecnologias apresentou mais de 15 mil instâncias, onde foi construído e gerado um modelo para prever as notas baixas, médias e altas do resultado do ENEM. O modelo atingiu uma acurácia de 71.9384 % das instâncias analisadas.

O trabalho de Santos et al. (2015), relatam o uso de indicadores, tais como: Insatisfeito, desanimado e nenhuma para realizar a validação cruzada dos dados e prever o quanto isso influencia os alunos em um ambiente virtual de ensino e futuramente contribuir para ações pedagógicas para apoiar estes alunos. Dessa forma, a interpretação do resultado minerado apresentou 72% de corretude na classificação dos dados com um coeficiente Kappa de 0,38, o que indica baixa concordância. A precisão da classificação ao considerar o desânimo, para esse método foi de 0.735, o que indica uma concordância considerável segundo o próprio autor. Considera-se uma etapa fundamental o retorno da interpretação e avaliação para prosseguimento de soluções eficazes.

Este trabalho, foca em receber, selecionar e processar os dados de três professores diferentes de uma mesma turma ao longo do Ensino Médio, assim identificando possíveis casos de recuperação e reprovação que necessitam de intervenções pedagógicas para melhorar o desempenho escolar na disciplina de física. O uso das árvores de decisão, aliada a explicação e o componente visual facilita a compreensão dos dados e anima o compartilhamento de futuras análises em outras disciplinas segundo Witten (2005).

4. Metodologia

Inicialmente, realizou-se uma compreensão do cenário onde os dados foram coletados e posteriormente utilizados. A partir dos dados obtidos, iniciou-se a etapa de desenvolvimento inspirado nos métodos KDD e EDM. KDD, significa a descoberta de conhecimento a partir dos dados, como podemos observar na Figura 4.1. que representa a linha de raciocínio aplicada juntamente com o uso da mineração de dados educacionais (EDM, Educational Data Mining) como parte do KDD, Han et al. (2006).

A Figura 4.1., condensa metodologia utilizada no artigo referente ao processo de descoberta de conhecimento (KDD): seleção dos dados, pré-processamento dos dados, transformação dos dados, mineração de dados educacionais (EDM), criação ou geração do experimento realizado e interpretação e avaliação dos resultados.

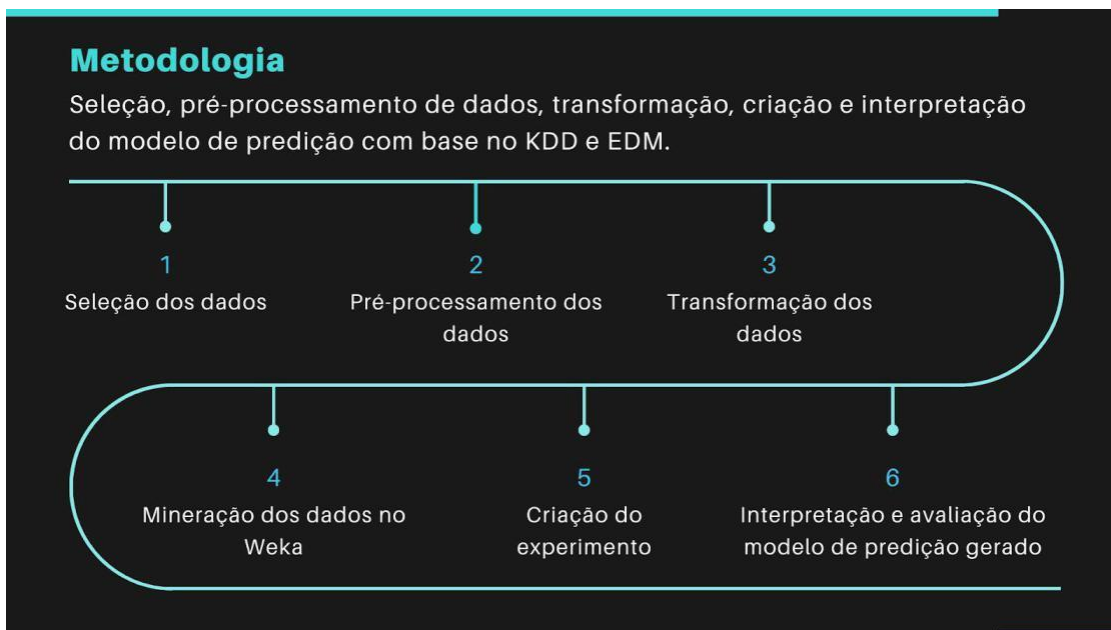


Figura 4.1. Processo utilizado na metodologia do artigo

Fonte: criado pelo autor

4.1. Contexto - Base de dados

Para a realização do trabalho, foram utilizados e analisados os dados coletados com três professores de física do Ensino Médio que disponibilizaram suas planilhas ao longo dos três anos do Ensino Médio de uma mesma turma. Os professores forneceram as planilhas referentes ao 1^a ano do EM de 2016, 2^a ano do EM de 2017 e do 3^a ano do EM de 2018. Cada professor de física do Ensino Médio leciona para 6 classes com 40 estudantes em média, resultando no estudo de aproximadamente 250 alunos durante 2016, 2017 e 2018.

A partir disso, os professores gostariam de compreender melhor se as provas, Provinhas, laboratórios, lições, desenvolvimento coletivo, faltas, participação em classe realmente os ajudam de forma efetiva na construção da média final dos estudantes. O objetivo final do trabalho é identificar o desempenho escolar dos alunos e consequentemente prever o porquê dos alunos ficarem de recuperação ou até mesmo reprovarem, já que os alunos têm diversos momentos para isto não acontecer.

Com base no exposto, surgem algumas perguntas de pesquisa: Como identificar as tendências dos alunos que ficam de recuperação? Como descobrir previamente o aluno que será reprovado?

4.2. Limpeza e Pré-processamento

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), a limpeza e pré-processamento dos dados engloba operações de limpeza para redesenhar os dados. Através desse passo, decidiu-se estratégias para superar informações não apresentadas ou até mesmo mudanças de sequência caso necessário.

Para a produção do trabalho, foram utilizados e analisados os dados fornecidos pela disciplina de física durante o Ensino Médio ao longo de 2016, 2017 e 2018. Após o recebimento das planilhas de cada professor, iniciou-se o pré-processamento dos dados, isso ocorre de forma sistêmica e aplicada para todo o conjunto de dados. Por exemplo: a Professora do 1º ano disponibilizou todos os dados em um único arquivo de Excel ".xlsx" e sabendo que o programa WEKA não tem capacidade para interpretar os dados corretamente, adota-se a estratégia de primeiramente remover, organizar e limpar acentuação, nomes longos, além de substituir o ";" por "," no arquivo ".csv" gerado a partir do ".xlsx" e ajustado dentro do programa Sublime Text a fim de padronizar o arquivo e consequentemente todas as outras planilhas recebidas.

Após essa etapa, inicia-se o processo de abertura do arquivo ".csv" no WEKA e posteriormente a criação do arquivo ".ARFF" (Attribute-Relation File Format), que descreve uma lista de instâncias que compartilham um conjunto de atributos com base na tabela ASCII (American Standard Code).

4.3. Transformação dos dados

A etapa de transformação de dados é considerada fundamental para organizar as variáveis e ressaltar os atributos considerados úteis em relação ao objetivo, ou seja, por meio da transformação é possível categorizar, reduzir, segmentar e condensar os dados de forma efetiva para alcançar os resultados projetados e esperados com base no algoritmo de mineração escolhido, Fayyad et al. (1996).

Para relembrar e facilitar o entendimento a frente do que vamos ver, veja a figura 4.2. abaixo com base no processo do KDD.

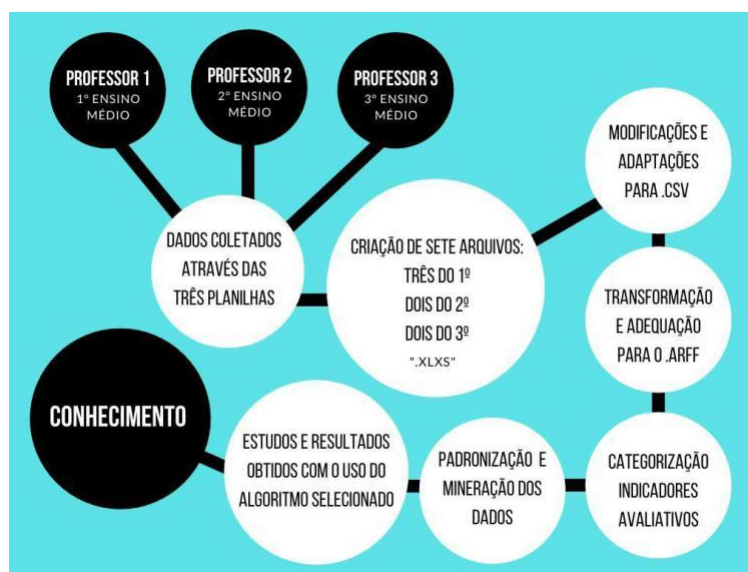


Figura 4.2. Overview dos passos inspirados no processo do KDD.

Fonte: criado pelo autor

No caso das planilhas recebidas, identificou-se a necessidade de geração de sete novas planilhas, sendo três planilhas referentes ao 1º EM e suas respectivas entregas de

conceito durante cada unidade. Já o 2º EM e o 3º EM são divididos em dois momentos, sendo um entrega de conceito no final do primeiro semestre e o outro momento no final do segundo semestre. Com isso em mente, dedica-se a transformação, adequação das informações em variáveis categorizadas para fácil compreensão e listadas a seguir.

Tabela 4.1. Categorização e organização

Séries	Variável	Métricas
1º EM indicadores avaliativos comuns nas três unidades	Prova 1	escala 0 a 10
	Prova 2	escala 0 a 10
	Plab (prova laboratório)	escala 0 a 10
	relatório (laboratório)	escala 0 a 10
	Des_coletivo (desempenho turma)	escala 0 a 10
	Part (participação)	escala 0 a 10
	Conceito	letras A, B, C e D
2º EM indicadores avaliativos comuns nos dois semestres.	Lab1 (laboratório)	escala 0 a 10
	Lab2 (laboratório)	escala 0 a 10
	Lab3 (laboratório)	escala 0 a 10
	P1 (prova 1)	escala 0 a 10
	P2 (prova 2)	escala 0 a 10
	CTSA (projeto interdisciplinar)	escala 0 a 10
	Conceito	letras A, B, C e D
3º EM indicadores avaliativos comuns nos dois semestres.	P1 (prova 1)	escala 0 a 10
	P2 (prova 2)	escala 0 a 10
	P3 (prova 3)	escala 0 a 10
	P4 (prova 4)	escala 0 a 10
	Provinhas	escala 0 a 10
	Lab (laboratório)	escala 0 a 10
	Lições	escala 0 a 10
	Autoaval (autoavaliação)	escala 0 a 10
	Faltas	escala 0 a 6
	Particip (participação)	escala 0 a 10
	1sem e 2sem (conceitos)	letras A, B, C e D

Observe que as variáveis escolhidas são instrumentos de avaliação de cada professor ao longo ano, desta forma, colaborando para padronizar minimamente e facilitar o entendimento dos dados transformados para obter uma melhor aplicação do algoritmo selecionado no minerador WEKA.

4.4. Mineração de dados

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a mineração de dados é o momento de exploração e seleção de modelo, escolhendo o algoritmo a ser utilizado para interpretar os dados anteriores já preparados.

A pergunta proposta a respeito do desempenho escolar e se é ou não possível prever se o estudante estará de recuperação nas primeiras unidades ou será reprovado durante o segundo semestre deve ser minerada a ponto de gerar uma análise e sugestões para o professor se precaver de tal acontecimento.

De acordo com Romero et al. (2008) e Witten (2005) o algoritmo J48 aplicado é normalmente utilizado no caso de classificação devido a efetividade e geração de

representatividade gráfica que podemos obter. A classificação nos ajuda a compreender e prever o desempenho escolar do estudante, o que vem a colaborar com o objetivo do trabalho. Conforme apresentado o algoritmo escolhido é o J48 por ser um dos algoritmos mais robustos e confiáveis em termos de mineração de dados e entendimento para as pessoas, devido a árvore de decisão criada e muitas vezes bem adequada ao problema analisado.

A Figura 4.3. mostra o exemplo da árvore de decisão criada no WEKA e que será melhor apresentada na próxima seção.

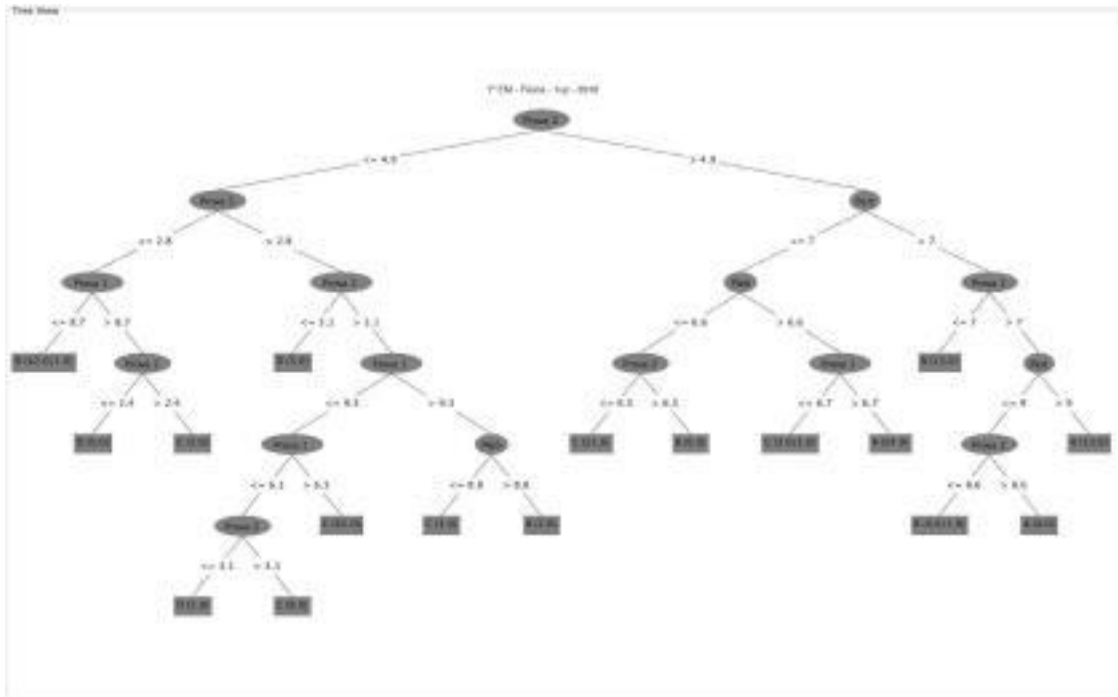


Figura 4.3 . Árvore de decisão 1º EM - Física - 1ª unidade 2016

Fonte: WEKA

5. Geração, avaliação e discussão

Essa etapa relata a interpretação dos dados minerados e extraídos de forma gráfica, numérica e também ramificada através da árvore de decisão que vimos anteriormente. Para os sete arquivos ".ARFF" preparados e criados para o uso no WEKA, estabeleceu-se por utilizar o método de validação cruzada, técnica selecionada para avaliar a capacidade propagação de um modelo, a partir de um conjunto de dados. No caso selecionado observa-se o número de conjuntos igual a 10, devido ao grande número de inputs disponíveis na base de dados. Importante notar que o mesmo algoritmo e parâmetro foi aplicado para a geração das sete árvores de decisão, variando apenas os inputs de um para o outro. Nenhum algoritmo retorna uma acurácia perfeita, porém todos os experimentos realizados apresentaram uma classificação satisfatória, veja o resumo nas figuras 5.1., 5.2. e 5.3.

=== Stratified cross-validation ===		=== Stratified cross-validation ===		=== Stratified cross-validation ===	
=== Summary ===		=== Summary ===		=== Summary ===	
Correctly Classified Instances	225	90 %	Correctly Classified Instances	217	87.1486 %
Incorrectly Classified Instances	25	10 %	Incorrectly Classified Instances	32	12.8514 %
Kappa statistic	0.8557		Kappa statistic	0.8225	
Mean absolute error	0.0448		Mean absolute error	0.0622	
Root mean squared error	0.1964		Root mean squared error	0.2147	
Relative absolute error	16.087 %		Relative absolute error	21.4297 %	
Root relative squared error	52.7382 %		Root relative squared error	56.4876 %	
Total Number of Instances	250		Total Number of Instances	249	
			Ignored Class Unknown Instances	1	
			Correctly Classified Instances	194	77.6 %
			Incorrectly Classified Instances	56	22.4 %
			Kappa statistic	0.6356	
			Mean absolute error	0.0972	
			Root mean squared error	0.2861	
			Relative absolute error	39.0328 %	
			Root relative squared error	81.3053 %	
			Total Number of Instances	250	

Figura 5.1. Validação cruzada estratificada 1º EM - Física - 1ª, 2ª e 3ª unidades de 2016

Fonte: WEKA

=== Stratified cross-validation ===		=== Stratified cross-validation ===		=== Stratified cross-validation ===	
=== Summary ===		=== Summary ===		=== Summary ===	
Correctly Classified Instances	160	67.7966 %	Correctly Classified Instances	177	75 %
Incorrectly Classified Instances	76	32.2034 %	Incorrectly Classified Instances	59	25 %
Kappa statistic	0.4458		Kappa statistic	0.6057	
Mean absolute error	0.1821		Mean absolute error	0.1172	
Root mean squared error	0.3777		Root mean squared error	0.3007	
Relative absolute error	61.2746 %		Relative absolute error	44.4517 %	
Root relative squared error	98.2026 %		Root relative squared error	83.0179 %	
Total Number of Instances	236		Total Number of Instances	236	

Figura 5.2. Validação cruzada estratificada 2º EM - Física - 1º e 2º semestre de 2017

Fonte: WEKA

=== Stratified cross-validation ===		=== Stratified cross-validation ===		=== Stratified cross-validation ===	
=== Summary ===		=== Summary ===		=== Summary ===	
Correctly Classified Instances	162	71.3656 %	Correctly Classified Instances	141	62.1145 %
Incorrectly Classified Instances	65	28.6344 %	Incorrectly Classified Instances	86	37.8855 %
Kappa statistic	0.5767		Kappa statistic	0.4224	
Mean absolute error	0.1517		Mean absolute error	0.2104	
Root mean squared error	0.3611		Root mean squared error	0.3772	
Relative absolute error	44.2112 %		Relative absolute error	62.179 %	
Root relative squared error	87.2416 %		Root relative squared error	91.7756 %	
Total Number of Instances	227		Total Number of Instances	227	

Figura 5.3. Validação cruzada estratificada 3º EM - Física - 1º e 2º semestre de 2018

Fonte: WEKA

A seguir apresenta-se a sequência de resultados obtidos através do algoritmo J48 e a criação da árvore de decisão para a segunda unidade e terceira unidade do 1º EM. Iniciamos com as figuras 5.4. e 5.5 para realizar uma melhor apreciação.

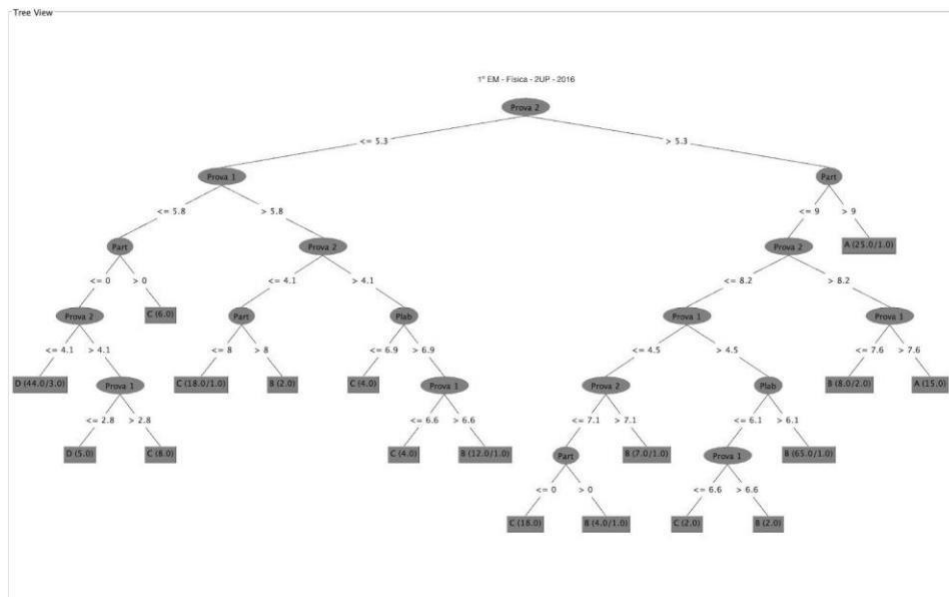


Figura 5.4. Árvore de decisão 1º EM - Física - 2ª unidade 2016

Fonte: WEKA

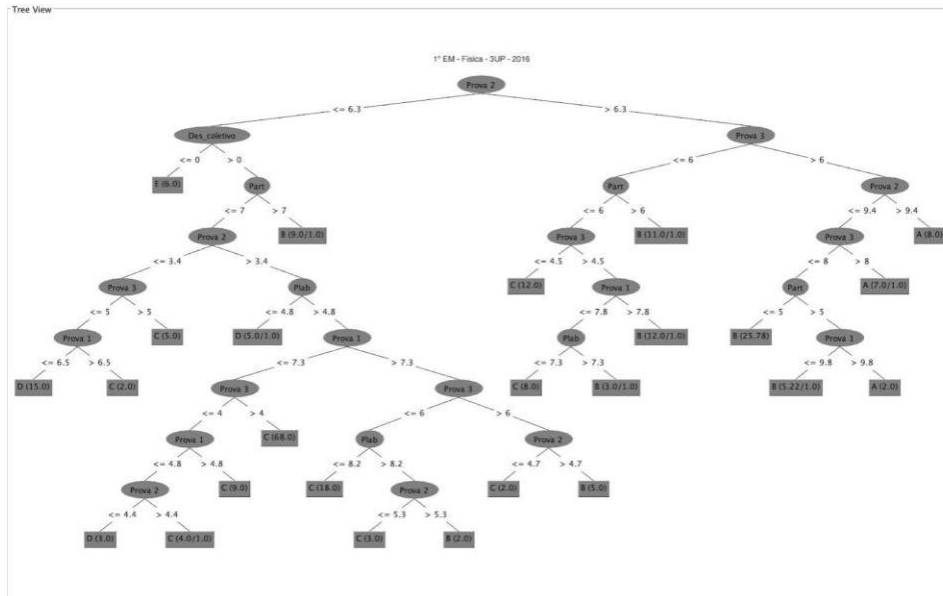


Figura 5.5. Árvore de decisão 1º EM - Física - 3ª unidade 2016

Fonte: WEKA

A partir da interpretação da árvore de decisões é possível compreender melhor os resultados. Inicia-se a seguir uma análise dos pontos mais interessantes encontrados pelo WEKA e que pode transpor para os próximos anos. Vamos analisar de forma completa a árvore de decisão a respeito do 1º EM de física da unidade 1 (1UP) do ano de 2016. Percebe-se que o algoritmo se baseia nas notas da Prova 2, figura 5.6., sendo que para o lado esquerdo as notas são abaixo ou igual a 4.9 e do lado direito superiores a 4.9. Com isto já podemos dizer que grande parte dos alunos que estão do lado esquerdo estão com conceitos "C" ou "D", sendo necessário a intervenção de forma pedagógica por parte do professor para alertar e colaborar para melhor desenvolvimento escolar e assim ocasionando a diminuição de estudantes de recuperação.

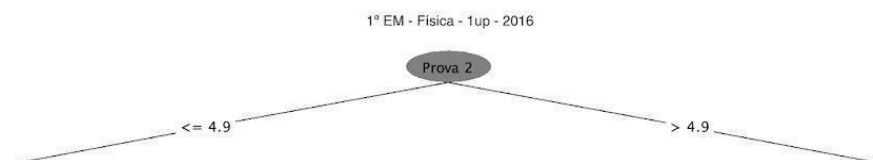


Figura 5.6. Root node - Prova 2 - 1º EM - Física - 1ª unidade 2016

Fonte: WEKA

Continuando a análise podemos perceber que todos os estudantes que foram acima de 4.9 na Prova 2 tem a tendência de ficar com conceito "A", "B" ou "C", ou seja, colaborando assim para uma primeira intervenção do professor para diminuir o número de alunos de recuperação. Observe a figura 5.7.

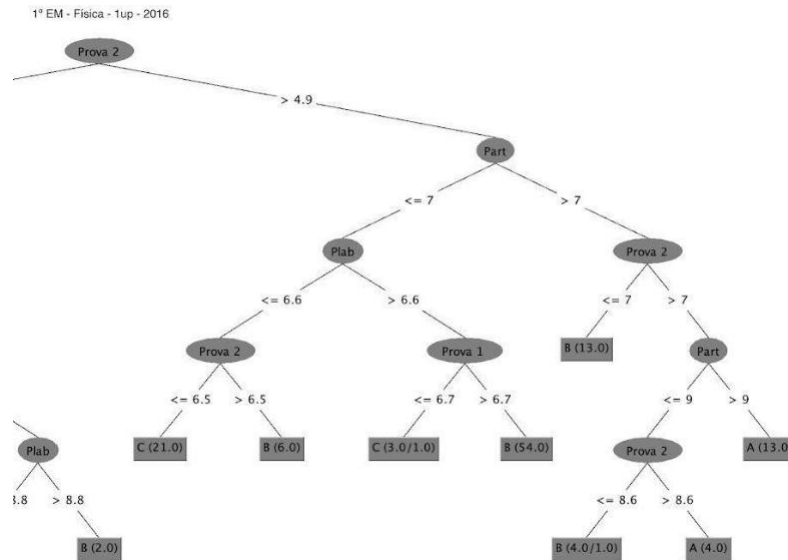


Figura 5.7. Prova 2 - lado direito - 1º EM - Física - 1ª unidade 2016

Fonte: WEKA

Sabe-se que o lado esquerdo da árvore de decisão alerta para os conceitos "C" e "D" e no caso da Prova 2 ser menor ou igual a 4.9, porém maior que 2.8, identificamos que os estudantes conseguem alcançar os conceitos "C" na maioria dos casos, salvo exceções em dois momentos com "D". Figura 5.8.

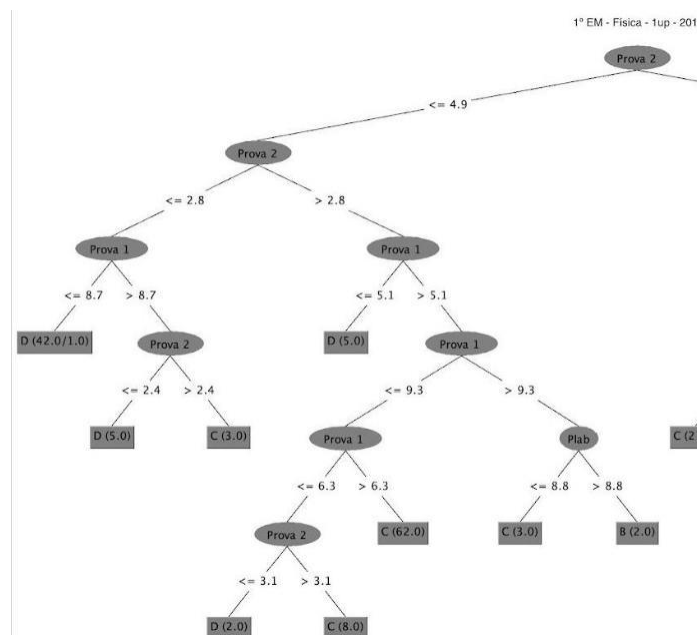


Figura 5.8. Prova 2 - lado esquerdo - 1º EM - Física - 1ª unidade 2016

Fonte: WEKA

Nota-se que o modelo escolhido, gerado e analisado, obteve uma árvore de decisão interessante que deve e pode ressaltar a criação de uma estratégia para que no

momento oportuno o professor venha a conversar com o estudante a respeito do desempenho escolar. Salienta-se aqui que as demais árvores de decisão também se inspiram na análise realizada e podem ser facilmente reproduzidas e interpretadas. Além da análise feita, sugere-se o envio do material minerado para que os professores envolvidos possam explorar e tirar novas conclusões.

Continua-se a análise para levantar novas possibilidades de intervenção em relação à terceira unidade do 1º EM, sendo nesse caso analisado as possíveis reprovações. Veja a Figura 5.9.

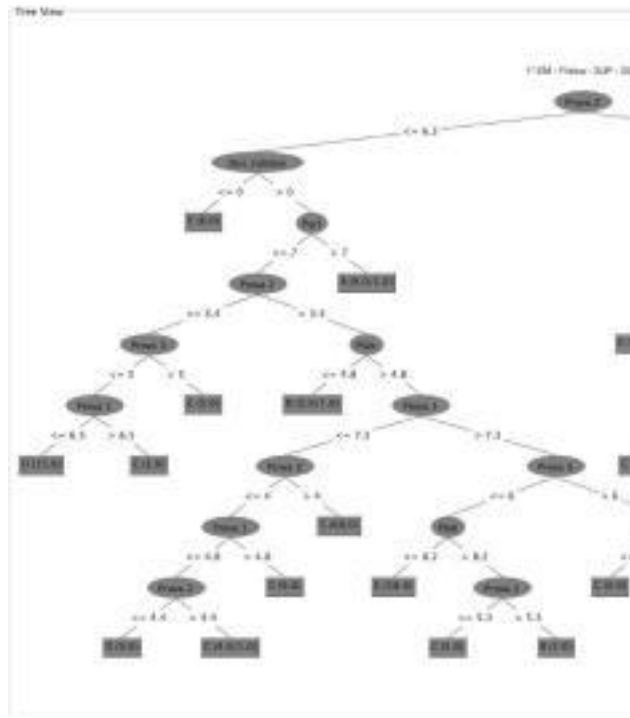


Figura 5.9. Prova 2 - lado esquerdo - 1º EM - Física - 3ª unidade 2016

Fonte: WEKA

Neste caso, podemos observar que o conceito "D" é visto em algumas combinações de resultados, tais como: Prova 2 menor do 6.3, Des_coletivo igual a 0, Part (participação) menor do 7.0 e Prova 2 menor do 3.4 concentra o número de 21 estudantes com provável chance de reprovação. Com isso em mente, é de extrema importância a conversa entre docente e discente para evitar o risco de reprovação.

A partir do do modelo criado e explanado com todas as etapas do KDD contempladas, pode-se aproveitar e realizar análises pontuais referentes ao 2º EM e o 3º EM a fim de colaborar para as estratégias adotadas por cada professor. A seguir as figuras 5.10. e 5.11. do 2º EM representadas pela árvore de decisão do primeiro e segundo semestre e mais adiante as figuras 5.12. e 5.13. do 3º EM durante o primeiro semestre e segundo semestre respectivamente.

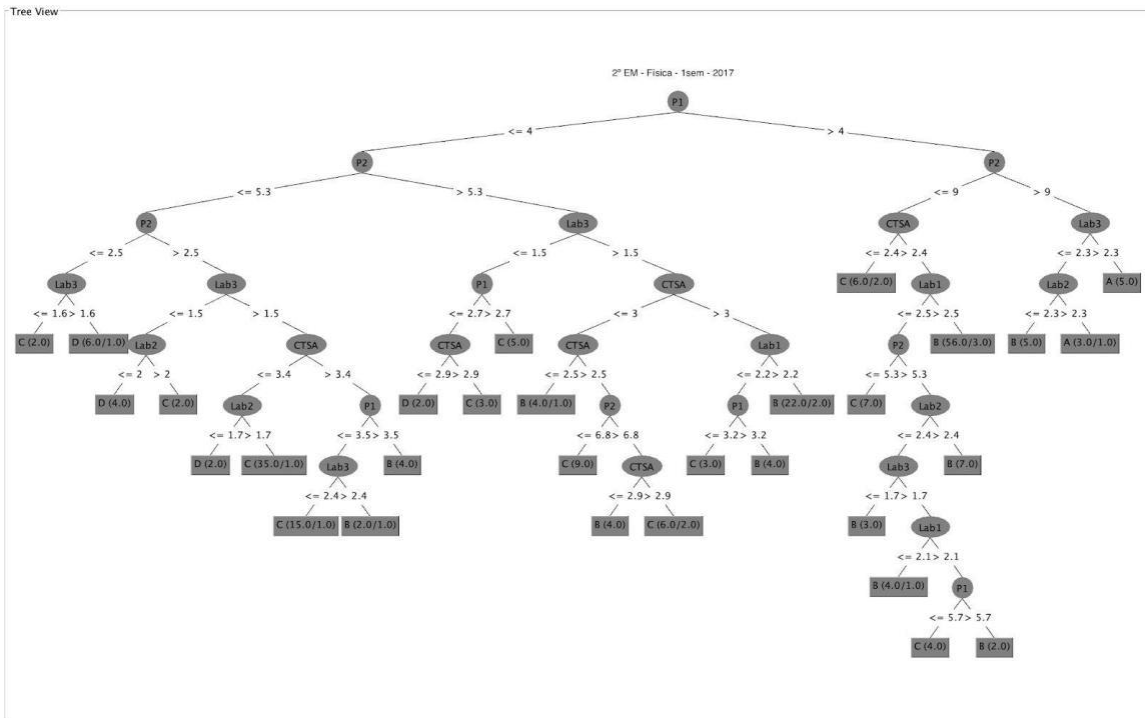


Figura 5.10. Árvore de decisão 2º EM - Física - 1º semestre de 2017

Fonte: WEKA

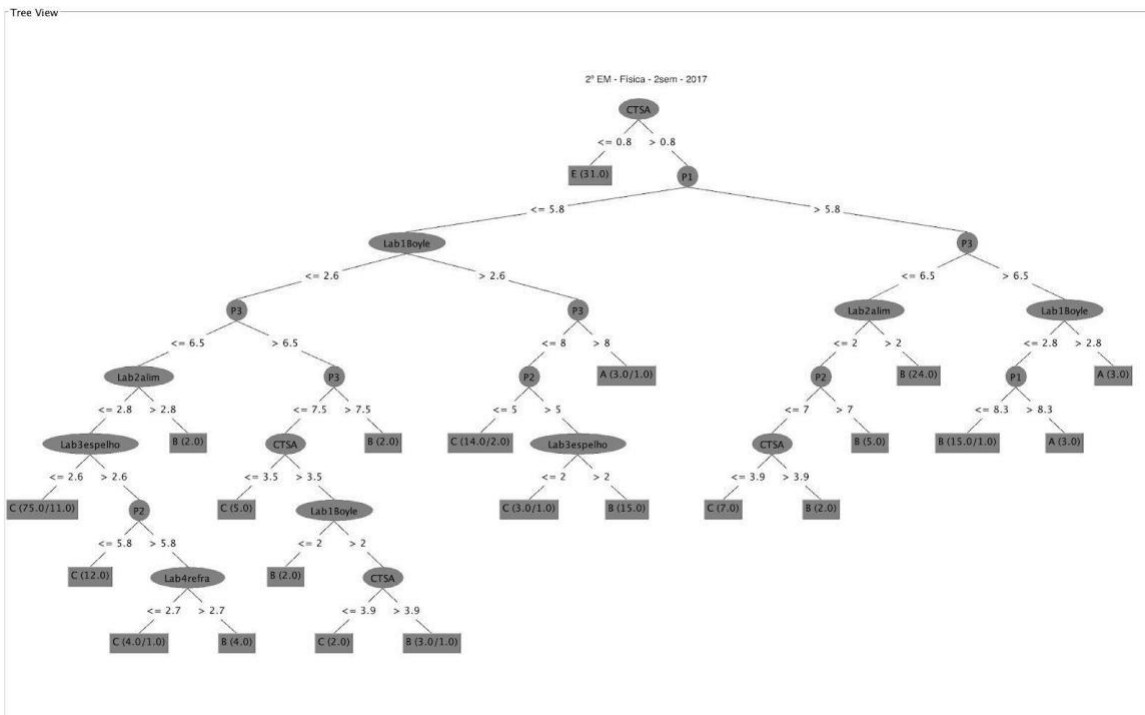


Figura 5.11. Árvore de decisão 2º EM - Física - 2º semestre de 2017

Fonte: WEKA

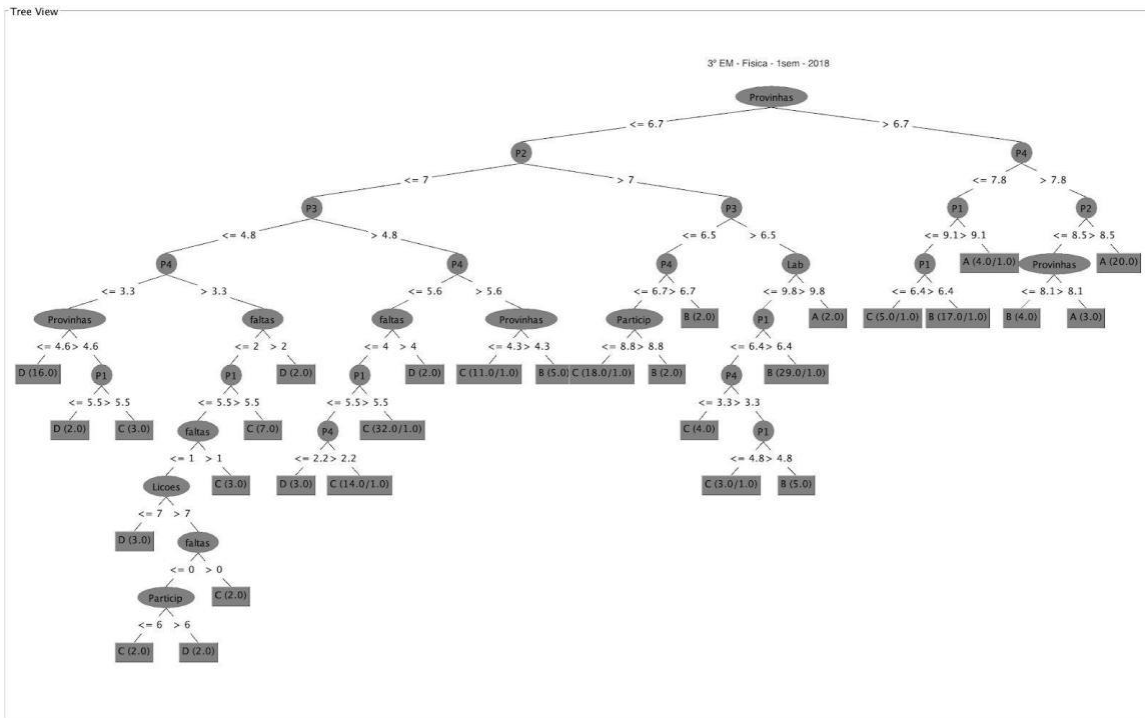


Figura 5.12. Árvore de decisão 3º EM - Física - 1º semestre de 2018

Fonte: WEKA

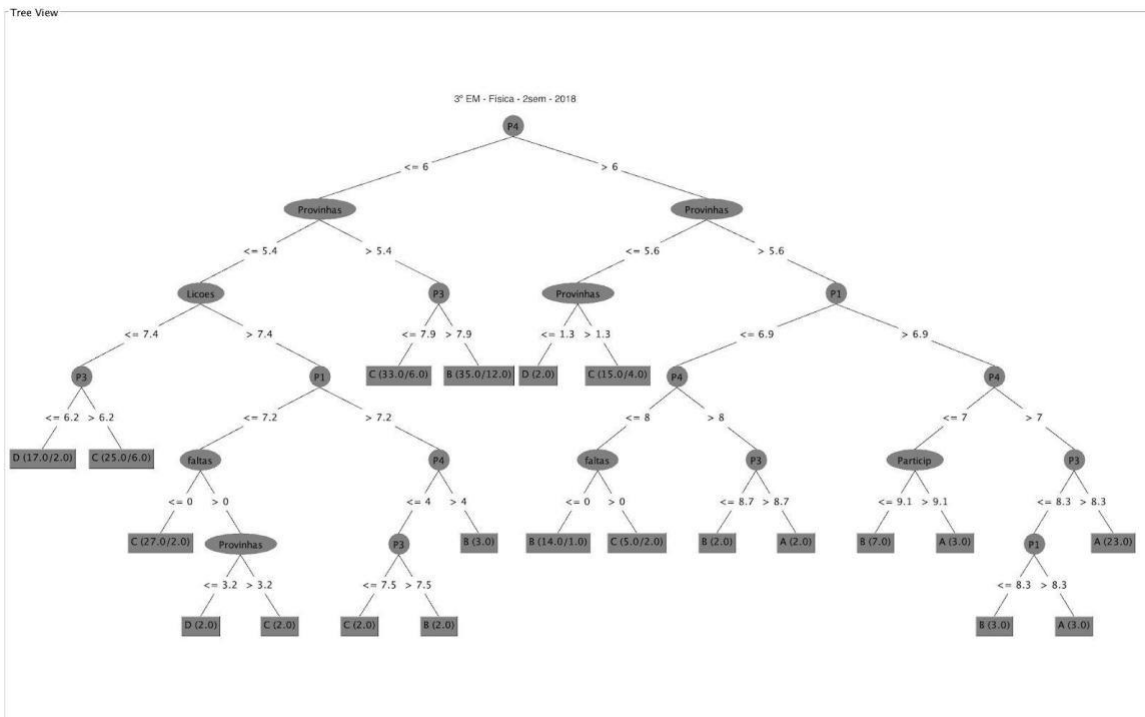


Figura 5.13. Árvore de decisão 3º EM - Física - 2º semestre de 2018

Fonte: WEKA

A árvore demonstra diversos resultados possíveis para o porquê de os estudantes tirarem o conceito "D" e ficar de recuperação ou até mesmo reprovarem. Podemos observar muito além do que foi dito acima, pois todos os dados desencadeiam para melhores apreciações.

6. Conclusão

Neste trabalho, foi possível evidenciar a importância da descoberta de conhecimento a partir do processo de KDD e EDM para gerar modelos que nos auxiliam ao longo do desempenho escolar na disciplina de física no Ensino Médio. Considerando o algoritmo J48 para classificar os dados, os professores das séries do Ensino Médio poderão sugerir atenção para pontos importantes e que não estavam claros ao observar apenas o arquivo ".xlsx", assim agindo e contribuindo com maior eficácia para melhorar o índice tanto de recuperandos, como estudantes com nota "D" nas primeiras unidades e semestre. Além disso, também é possível estabelecer notas medianas caso os estudantes tenham algum descuido. Neste mesmo intuito, pode-se perceber também as chances de reprovação dos estudantes durante a última unidade e segundo semestre a fim de garantir maior atenção em alguns pontos para o discente não ser reprovado.

Por fim, vale ressaltar que os inputs escolhidos fazem diferença para o sucesso da predição e interpretação dos resultados obtidos, assim como a escolha da árvore de decisão para fácil compreensão geral. Para trabalhos futuros vale debater qual o espaço da mineração de dados dentro da escola e voltar o questionamento de como os dados podem influenciar no poder de decisões futuras em diversos aspectos pedagógicos.

Este trabalho defendeu o uso da metodologia do KDD e EDM aliados ao desempenho escolar desde o momento dos dados brutos até o refinamento do mesmo. Cabe à escola, incorporar e explorar ao máximo os programas e criar as bases para o bom funcionamento e automatização de mineração de dados colhidos dentro da escola.

7. Referências

- Alves, R., Cechinel, C., Queiroga E., (2018). Predição do desempenho de Matemática e Suas Tecnologias do ENEM utilizando técnicas de Mineração De Dados. VII Congresso Brasileiro de Informática na Educação. CBIE.
- Baker, R.S.J.d. (2010). Data Mining for Education. McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.
- Baker, R.S.J.d., I. S. d. C. A. (2011). Mineração de dados educacionais: Jornada de Atualização em Informática na Educação - JAIE 2012 25/29 Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, 19(2).
- Han, J., Kamber, M., (2006), Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, Second Edition.
- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic. (1996). From Data Mining to Knowledge Discovery in Databases. Aai Press, p.37-54.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology & Society, 17(4), 49-64

Rabelo, H., Burlamaqui A., Valentim, R., Rabelo, D., Medeiros, S. (2017). Utilização de técnicas de Mineração de Dados Educacionais para a predição de desempenho de alunos de EaD em Ambientes Virtuais de Aprendizagem. VI Congresso Brasileiro de Informática na Educação. CBIE.

Romero, C., Ventura, S., Espejo, P.G., Hervas, C. (2008). Data Mining Algorithms to Classify Students. In Proceedings of the International Conference on Educational Data Mining, 8-17.

Santos, F., Bercht, M., Wives L., (2015). Evidência do desânimo de alunos em um ambiente virtual de ensino e aprendizagem: uma proposta a partir da mineração de dados educacionais. CINTED-UFRGS, Novas Tecnologias na Educação.

Vosgerau, R. (2014). Estudos de revisão: implicações conceituais e metodológicas. Revista Diálogo Educacional 14(41): 165-189.

Witten, I. H., Frank, E. (2005). “Data Mining: Practical Machine Learning Tools and Techniques”. 2nd edition, Morgan Kaufmann Publishers, San Francisco, CA.

Witten, I. H, Curso online Data Mining with Weka. Disponível em:
<<http://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/>> Acesso em: 29 novembro de 2020.