

Replicação de [Jyoti et al. 2015]

Jean B. Teixeira¹

¹Doutorando em Ciência da Computação – Universidade Federal de Campina Grande (UFCG)

jean@copin.ufcg.edu.br

Resumo. O objetivo desse trabalho é reproduzir o experimento, do artigo aqui apresentado, utilizando o que for pertinente do conteúdo apresentado ao longo do semestre na disciplina de Fundamentos de Pesquisa em Ciência da Computação II pelo professor Nazareno Ferreira.

1. Estudo original

O artigo, objeto de estudo dessa replicação, foi publicado em 2015 na International Conference on Futuristic trend in Computational Analysis and Knowledge Management com o objetivo de analisar o melhor classificador para dados de filmes com base nas avaliações dos usuários e, em seguida, a classificação é usada para fazer as recomendações para os usuários.

O trabalho apresenta a problemática do grande volume de dados que fazem parte do nosso cotidiano e a dificuldade de realizar escolhas de itens relevantes em um enorme espaço de opções possíveis. Para isso, surgem os sistemas de recomendação, que estimam a preferência dos usuários sobre os itens disponíveis, nesse caso filmes, e os recomendam de maneira proativa listando aqueles que sejam de seu interesse [Ricci et al. 2015]. As avaliações do usuário são usadas para classificar dados em várias categorias que podem ser úteis para gerar essas recomendações.

De forma geral, classificação é o processo de atribuir um rótulo de um conjunto predefinido a um objeto, por exemplo, classificando um filme de acordo com seu gênero ou avaliação do usuário. Entretanto, no contexto de mineração de dados, a classificação consiste em uma análise de dados em duas etapas, a primeira onde um modelo é gerado a partir de um conjunto de dados de treinamento usando algoritmos estatísticos e de aprendizado de máquina capazes de, em uma segunda etapa, prever qual a classe um novo dado até então desconhecido pertence [Han et al. 2011]

Para realizar as classificações os autores utilizaram-se da ferramenta de mineração de dados WEKA¹ e um conjunto de dados de filmes disponíveis na internet. Eles apresentam uma comparação entre diferentes métodos de classificação, aplicados aos dados de filmes, analisando algumas métricas² de avaliação como precisão, recall, taxa de erro entre outras. A avaliação de um classificador é realizada a partir da comparação entre as classes preditas pelo modelo e as classes verdadeiras dos exemplos. Todas as métricas tem um foco comum que é medir o quanto distante o modelo está da classificação perfeita, porém cada uma delas tem suas particularidades.

¹Saiba mais em: <https://ieeexplore.ieee.org/document/7155014>

²Saiba mais em: <https://diegomariano.com/metricas-de-avaliacao>

O artigo também apresenta alguns trabalhos relacionados e ressalta a importância da classificação e apresenta alguns algoritmos³ que já foram usados para gerar recomendações, por exemplo, classificadores bayesianos, KNN e árvores de decisão.

Link para o artigo: <https://ieeexplore.ieee.org/document/7155014>

2. Objeto da reprodução

Para a reprodução iremos utilizar o que foi especificado na seção V do artigo, já especificado anteriormente, com exceção da subseção D que explana sobre a geração da recomendação a partir dos classificadores, ou seja, o foco é fazer a análise exploratória dos dados, aplicar alguns algoritmos de classificação e compará-los de acordo com algumas métricas escolhidas. A tabela 1, mostra o resultado final (média das métricas em relação as cinco classes) obtido pelos autores e deixa claro todas essas configurações, classificadores e métricas, que foram citadas acima:

Tabela 1. Resultados dos classificadores (média entre os resultados das classes)

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	MAE	MSE
ZeroR	0.342	0.342	0.117	0.342	0.174	0.5	0.2992	0.3868
Naive Bayes	1	0	1	1	1	1	0.009	0.0015
J48	1	0	1	1	1	1	0.0	0.0

Após apresentarem os resultados, os autores concluíram que o classificador ZeroR fornece classificações imprecisas para dados de filmes, e por isso, suas saídas não podem ser utilizadas para fazer recomendações. Já os classificadores J48 e bayesiano trazem bons resultados mas o J48 consegue superá-lo devido sua taxa de erro baixa e alta especificidade.

Sendo assim, após a análise dos resultados, os pesquisadores afirmam que o algoritmo de árvore de decisão J48 é o que melhor se adapta aos dados dos filmes e o seu resultado pode ser utilizado para realizar recomendações aos usuários auxiliando-os a um fácil acesso aos itens de seus interesses em meio a uma enorme quantidade de opções de escolhas. Vale ressaltar, que eles também afirmam que o domínio pode ser ampliado atendendo não apenas a dados de filmes mas sites de compras e redes sociais por exemplo.

3. Metodologia original

3.1. Coleta de dados

O conjunto de dados utilizado é denominado MovieLens que é coletado e disponibilizado pela Group Lens Research⁴. O conjunto possui cerca de 100.000 mil classificações realizadas por 943 usuários em 1.682 filmes e foram utilizados três atributos, o id do usuário, id do filme e classificação do filme feita pelo usuário, essa classificação é definida em cinco classes, ou seja, ruim, ok, médio, bom e excelente.

³Saiba mais em: <https://ilumeo.com.br/todos-posts/2020/06/22/um-tour-pelos-10-principais-algoritmos-de-machine-learning>

⁴Saiba mais em: <https://grouplens.org/datasets/movielens/>

Foram realizados alguns pré-processamentos nesse conjunto de dados para facilitar sua utilização durante o processo de classificação, por exemplo, o atributo de classificação dos filmes, que é a variável alvo, foi convertida de categórica para numérica: 1=ruim, 2=ok, 3=médio, 4=bom e 5=excelente.

Os classificadores utilizados na pesquisa foram, ZeroR, Naive Bayes e J48, e as métricas utilizadas para avaliar o desempenho deles foram, True Positive Rate, False Positive Rate, Precision, Recall, F-Measure, ROC Area, Mean Absolute Error (MSE) e Mean Square Error (MSE).

3.2. Dados gerados

Após a coleta dos dados e a etapa de pré-processamento, os autores obtiveram um novo conjunto de dados e os mesmos foram submetidos aos três classificadores escolhidos onde podemos observar algumas previsões, conforme podemos observar na Tabela 2.

Tabela 2. Amostra do resultado das classificações dos algoritmos

Classifier	user_id	movie_id	predicted_rating	real_rating
ZeroR	1	1245	4	3
Naive Bayes	22	452342	4	4
J48	250	5452	1	1
Naive Bayes	472	11257	3	2
ZeroR	561	42104	2	1

3.3. Análise de dados

Por fim, com os dados prontos para um melhor uso foram feitas algumas análises das variáveis, na figura 1 é exibida a distribuição dos dados em relação as cinco classes na seguinte ordem, ruim, ok, médio, bom e excelente. Podemos perceber que existe a predominância da classe bom (verde escuro), seguido de médio (azul claro) enquanto a classe com menor ocorrência é a ruim (azul escuro).

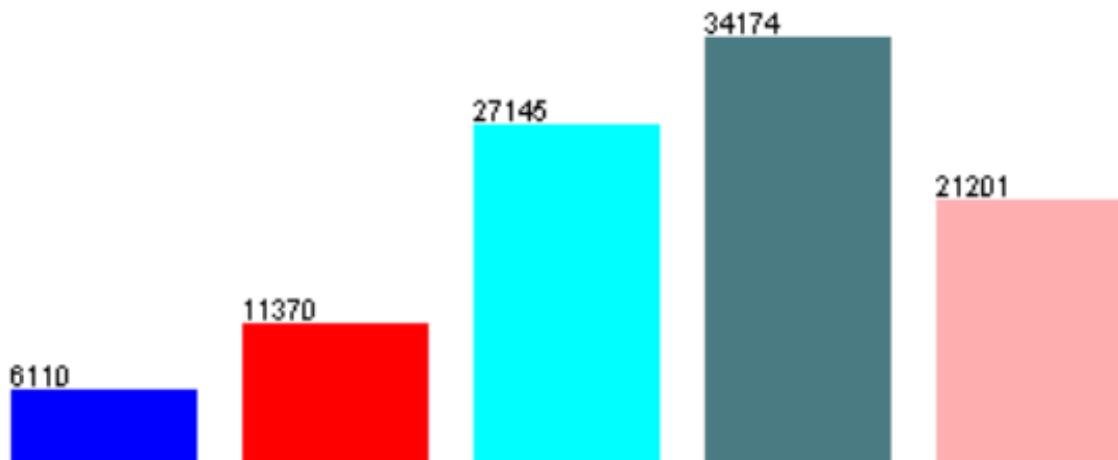


Figura 1. Distribuição das classes.

O mesmo foi feito para os outros dois atributos, user_id e movie_id conforme as imagens 2 e 3.

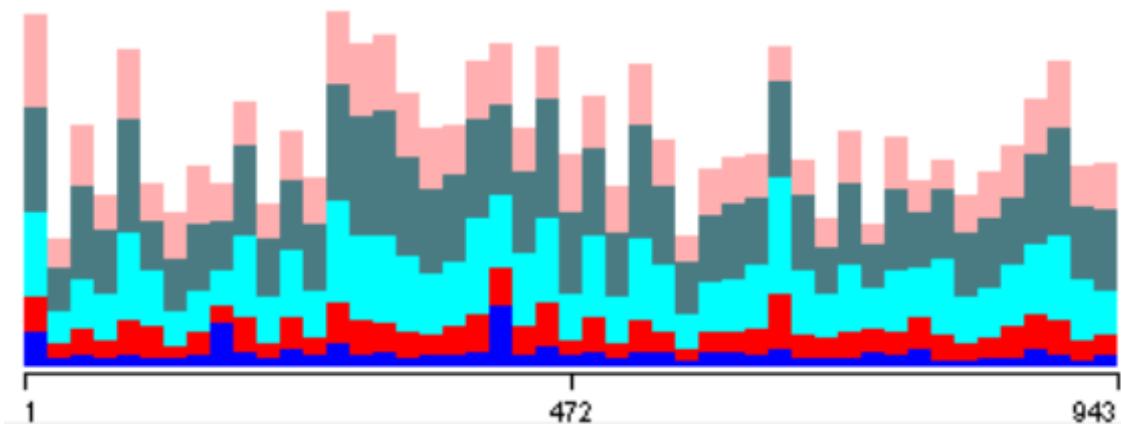


Figura 2. Distribuição das classes em relação a variável user_id.

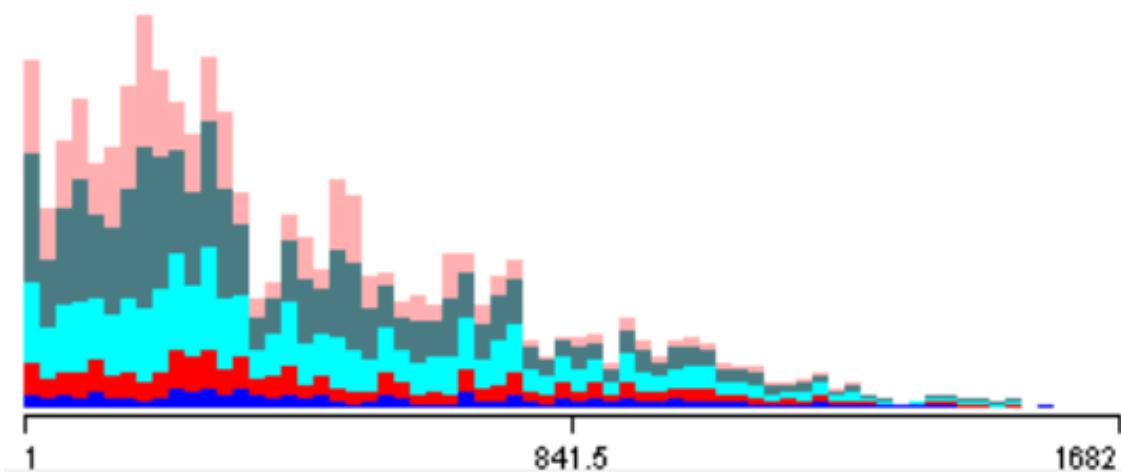


Figura 3. Distribuição das classes em relação a variável movie_id.

Após a análise, os dados foram submetidos a cada algoritmo e foram obtidos resultados para cada classe em cada um deles, como exemplo, veremos o resultado do algoritmo ZeroR na Tabela 3, pois o resultado final comparando todos os algoritmos já foi apresentado na tabela 1.

Tabela 3. Resultado do classificador ZeroR

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Classe
0	0	0	0	0	0.5	Ruim
0	0	0	0	0	0.5	Ok
0	0	0	0	0	0.5	Médio
1	1	0.342	1	0.509	0.5	Bom
0	0	0	0	0	0.5	Excelente
0.342	0.342	0.117	0.342	0.174	0.5	Média

4. Diferenças metodológicas com o estudo original

Para a reprodução do estudo, optamos por utilizar uma nova base de dados de filmes chamada de IMDb⁵ e além de uma base diferente utilizaremos mais variáveis como por exemplo, gênero do filme, ano de lançamento, idioma entre outras. O objetivo principal continua sendo qual o melhor algoritmo para realizar classificações de avaliações de filmes mas além disso, visto que utilizaremos mais variáveis, quais delas são mais relevantes para o classificador.

Outra mudança será em relação a ferramenta, no trabalho original foi utilizado a ferramenta WEKA, nós iremos utilizar tanto a ferramenta R, essa mais voltada para a análise exploratória dos dados e inferências, como o python para o desenvolvimento dos modelos de classificação, avaliação das métricas e variáveis importantes. Sobre as métricas, serão utilizadas algumas das que foram utilizadas no artigo original mas com uma apresentação um pouco diferente e sobre os classificadores, utilizaremos o Naive Bayes, Random Forest e XGBoost. Apesar das mudanças, a metodologia seguida será a mesma, teremos a etapa de coleta dos dados que estão disponíveis na internet, teremos uma análise exploratória das variáveis utilizadas tanto em sua forma bruta quanto utilizando sumarizações e alguns gráficos para facilitar a interpretação, caso necessário será realizado um pré-processamento para tornar os dados mais interpretáveis para os algoritmos, serão desenvolvidos os modelos de classificação e por último os resultados analisados, serão apresentados e discutidos.

Outra análise que será realizada é que com o melhor algoritmo iremos comparar, através de inferência estatística, o efeito que as variáveis mais preditoras tem sobre esse classificador. Iremos comparar a média da diferença entre as classificações com e sem as variáveis mais preditoras, ou seja, aquelas mais relevantes para a classificação.

Em resumo, as contribuições dessa replicação serão:

- Indicar o melhor algoritmo, dentre os escolhidos, tem maior desempenho para classificar as avaliações dos usuários sobre filmes;
- Apresentar as variáveis mais preditoras;
- Medir se existe efeito relevante no desempenho do melhor algoritmo, com e sem essas variáveis mais preditoras;

5. Resultado da replicação

Nesta seção, iremos apresentar o resultado do experimento realizado conforme foi previamente descrito anteriormente. Teremos a análise exploratória de algumas variáveis, o resultado da aplicação de três algoritmos de aprendizagem de máquina e algumas conclusões mediante os resultados alcançados.

5.1. Análise Exploratória dos Dados

Para uma melhor compreensão, na Tabela 4 temos o dicionário de dados da base utilizada no experimento. De forma empírica, os autores escolheram algumas variáveis que, teoricamente, podem ter certa influência na variável alvo, nesse caso a variável weighted_average_vote que é a classificação média ponderada total do filme.

⁵Saiba mais em: <https://www.imdb.com/interfaces/>

Tabela 4. Dicionário dos Dados

Atributo	Tipo	Descrição
imdb_title_id	chr	ID do filme no IMDb
year	int	Ano de lançamento
duration	int	Duração
reviews_from_users	dbl	Número de comentários de usuários
reviews_from_critics	dbl	Número de comentários dos críticos
total_votes	dbl	Total de votos recebidos
males_allages_avg_vote	dbl	Classificação média de todos os usuários do sexo masculino (todas as idades)
males_allages_votes	dbl	Número de votos de todos os usuários do sexo masculino (todas as idades)
females_allages_avg_vote	dbl	Classificação média de todos os usuários do sexo feminino (todas as idades)
females_allages_votes	dbl	Número de votos de todos os usuários do sexo feminino (todas as idades)
us_voters_rating	dbl	Classificação média de usuários americanos
us_voters_votes	dbl	Número de votos de usuários americanos
non_us_voters_rating	dbl	Classificação média de usuários não americanos
non_us_voters_votes	dbl	Número de votos de usuários não americanos
weighted_average_vote	dbl	Classificação média ponderada total

A seguir, a Figura 4, apresenta a distribuição da variável alvo (weighted_average_vote). Aparentemente segue uma distribuição normal mas contém uma cauda longa para a esquerda e algumas ocorrências de dados nas extremidades direita quase próximo a classificação 10. Vale ressaltar que a maior concentração dos dados está entre a classificação 5,0 e 7,5.

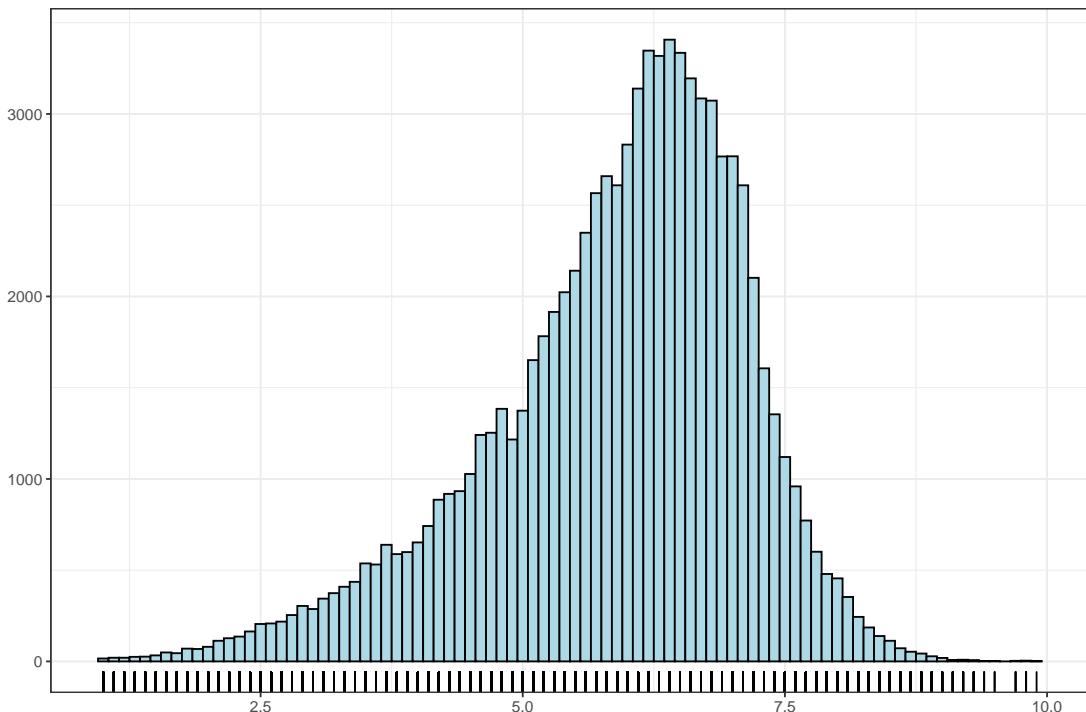


Figura 4. Distribuição da Classificação Média Ponderada Total.

Na Tabela 5 podemos conferir que poucos filmes possuem uma classificação média maior/igual a 9,5 e que a maioria foram lançados nos últimos 10 anos. Talvez isso possa indicar que a qualidade dos filmes tiveram uma melhora nesse período.

Tabela 5. Filmes com classificação maior/igual a 9,5.

Título	Classificação	Ano	Duração
Suvarna Sundari	9.9	2019	120
Lejos de Casa pelicula Venezolana	9.8	2020	87
Jeeudo	9.8	2020	150
Ek	9.8	2019	129
Notuku Potu	9.8	2017	121
Meka Suri	9.7	2020	86
Hopeful Notes	9.7	2010	94
Vaarthakal Ithuvare	9.7	2019	121
The Moving on Phase	9.5	2020	85
Isha	9.5	2020	118

Em relação a distribuição da duração dos filmes, ao observar a Figura 5, percebemos que a grande maioria está dentro do que costuma-se ver que seria entre uma hora até cerca de duas horas e meia. Mas, existe uma cauda longa para a extremidade direita indicando filmes que ultrapassam 800 minutos, ou seja, são mais de 13 horas de filme.

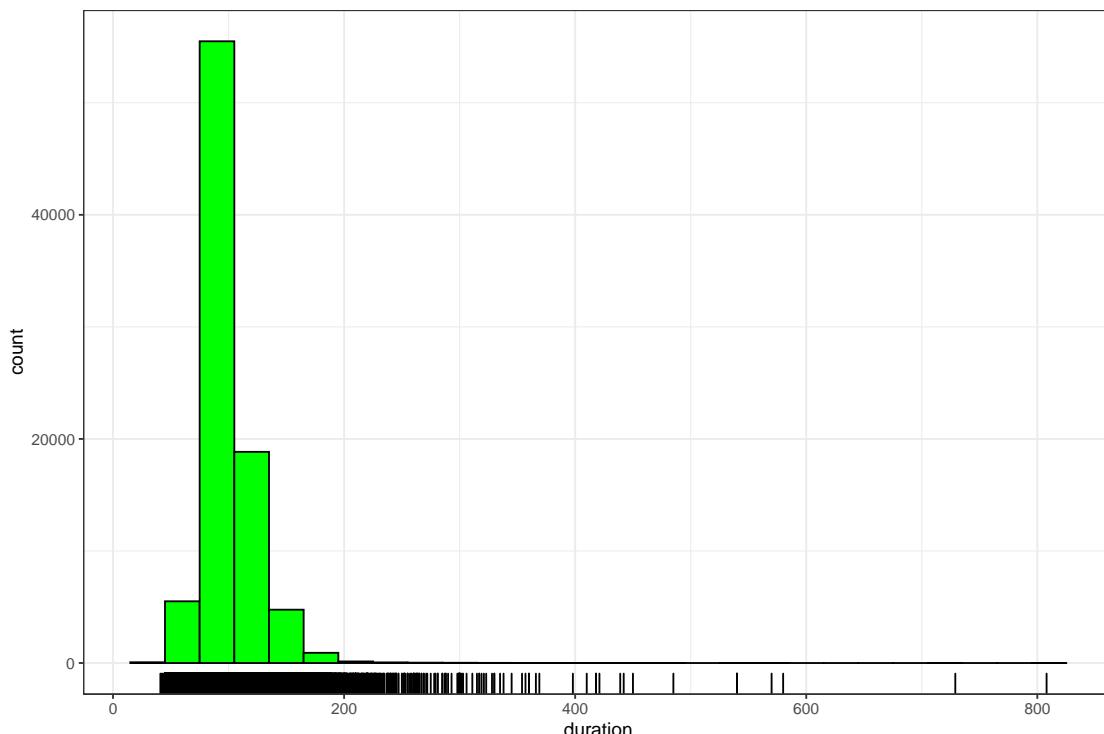


Figura 5. Distribuição da Duração dos Filmes.

Na Tabela 6, podemos identificar os principais filmes com maior tempo de duração que fazem parte da extremidade direita do gráfico anterior, curioso que o maior é recente com lançamento em 2016.

Tabela 6. Filmes com duração maior que 300 minutos.

Título	Duração (min)	Classificação	Ano
La flor	808	7.7	2016
Out 1, noli me tangere	729	7.8	1971
Khleb - imya sushchestvitelnoe	580	5.8	1988
Orbius	570	9.3	2020
Ebolusyon ng isang pamilyang Pilipino	540	7.8	2004
Heremias: Libro primo	540	7.0	2006
Kagadanan sa banwaan ning mga engkanto	540	7.4	2007
Hele sa hiwagang hapis	485	7.4	2016
Melancholia	450	7.4	2008
Hitler, un film dalla Germania	442	7.6	1977

Agora, vamos observar na Figura 6, a distribuição dos dados em relação ao ano de lançamento dos filmes. Podemos perceber a evolução que houve na produção de filmes que foi crescendo de forma exponencial, principalmente a partir dos anos 2000.

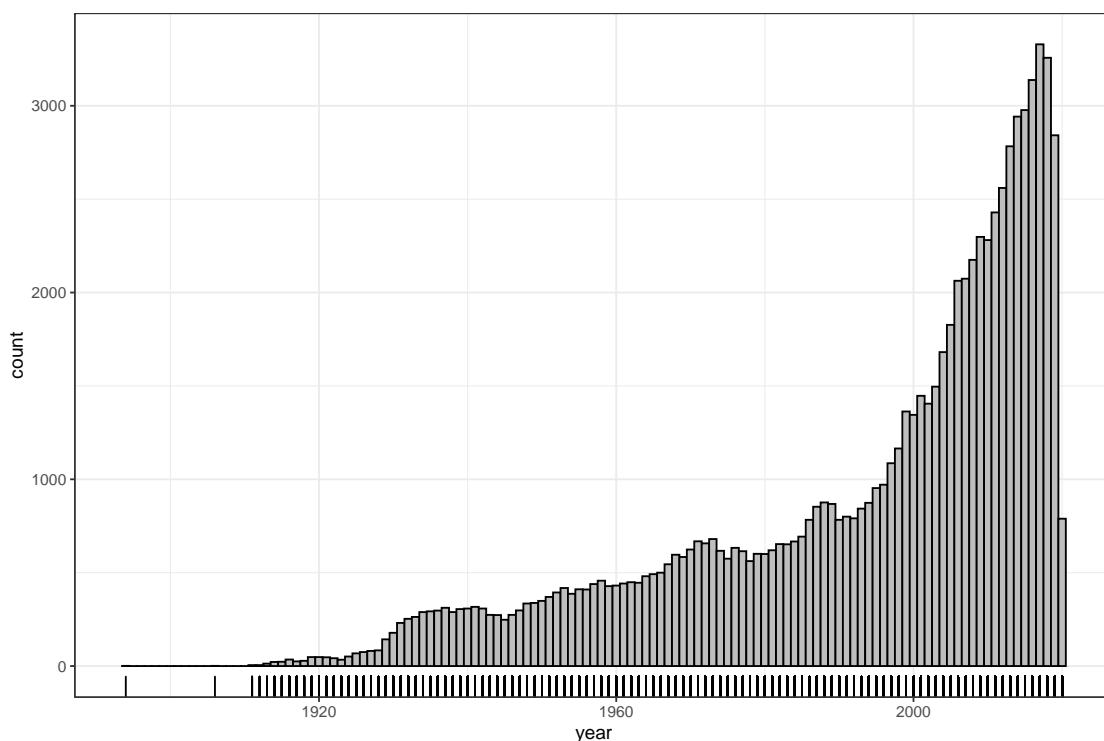


Figura 6. Distribuição dos Filmes por Ano.

A Tabela 7, mostra os 10 filmes mais antigos da nossa base de dados, com ano de lançamento anterior a 1915.

Tabela 7. Top dez filmes mais antigos.

Título	Ano	Duração	Classificação
Miss Jerry	1894	45	5.9
The Story of the Kelly Gang	1906	70	6.1
Den sorte drøm	1911	53	5.8
L’Inferno	1911	68	7.0
Oborona Sevastopolya	1911	100	6.0
Pinocchio	1911	50	5.7
Karadjordje	1911	80	6.2
Cleopatra	1912	100	5.2
From the Manger to the Cross; or, Jesus of Nazareth	1912	60	5.7
Independenta Romaniei	1912	120	6.7

Agora, iremos verificar se existe alguma correlação entre a variável alvo, classificação média ponderada total, e alguma outra variável da base. A escolha será de forma empírica, com base no que os autores acreditam que possa haver uma possível correlação, por exemplo, iremos verificar as variáveis: total_votes, males_allages_avg_vote e non_us_voters_rating.

Na Figura 7, observamos o gráfico de dispersão entre a variável total de votos e a classificação média dos filmes. Podemos perceber que não existe uma correlação linear entre as variáveis, o coeficiente de correlação de Pearson foi de 0,18.

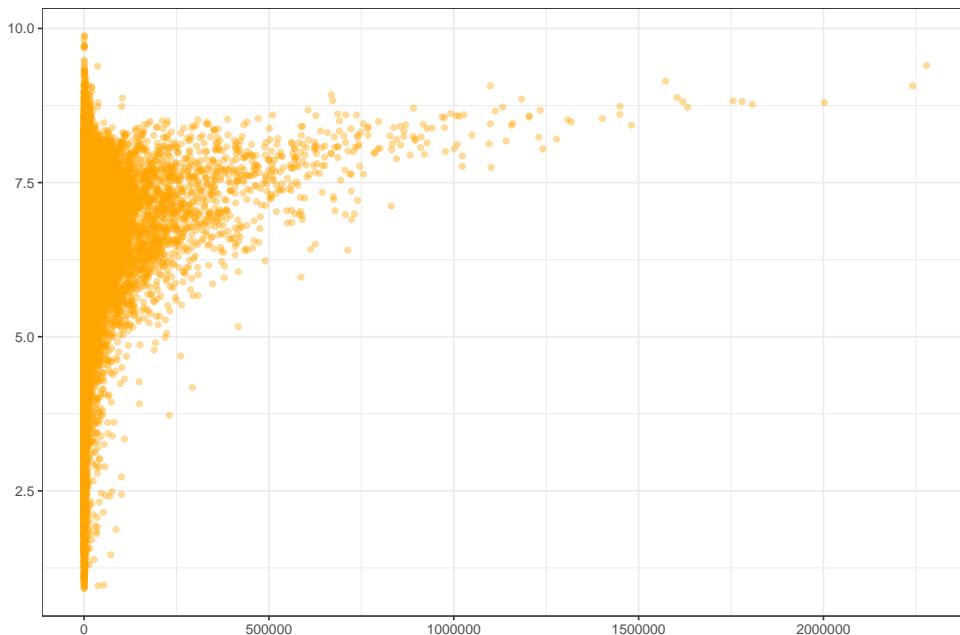


Figura 7. Total de votos X Classificação média ponderada total.

Já na Figura 8, podemos perceber que existe uma correlação positiva e forte entre as variáveis com um coeficiente de correlação de Pearson igual a 0,98.

Nesse caso apresentado pela Figura 9, também é perceptível uma correlação considerável e positiva entre as variáveis analisadas com um coeficiente de correlação de

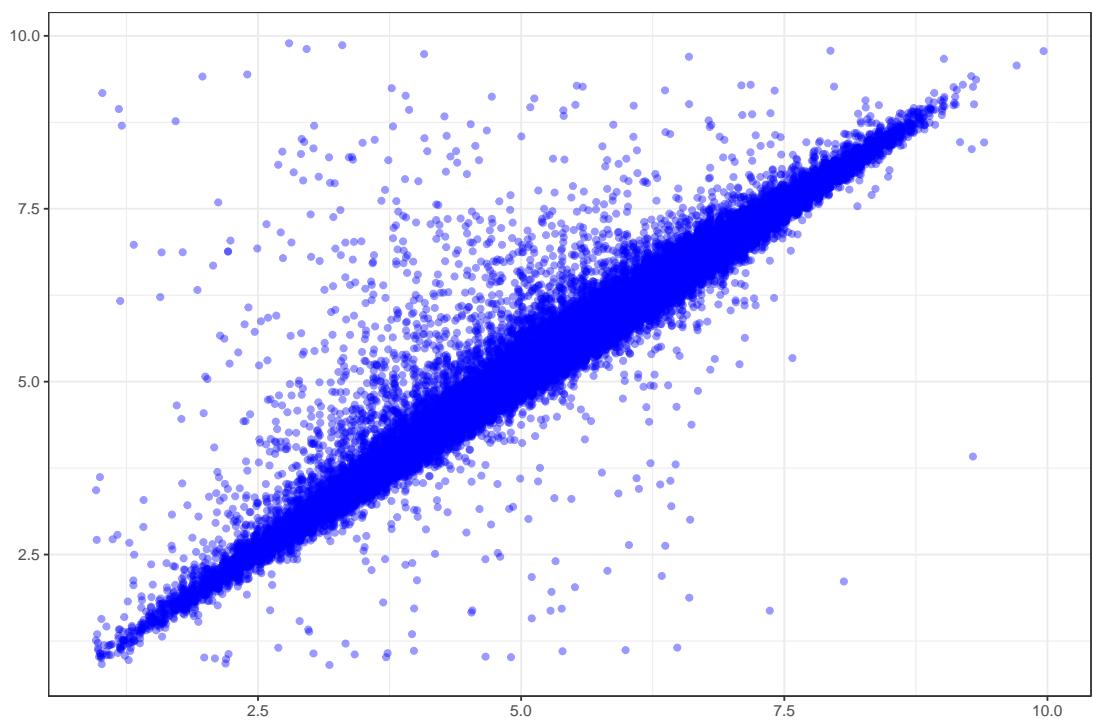


Figura 8. Classificação média dos homens X Classificação média ponderada total.

Pearson igual a 0,96.

Provavelmente essas variáveis com uma correlação positiva e considerável terão uma influência maior nos modelos preditivos, verificaremos essa hipótese na próxima seção.

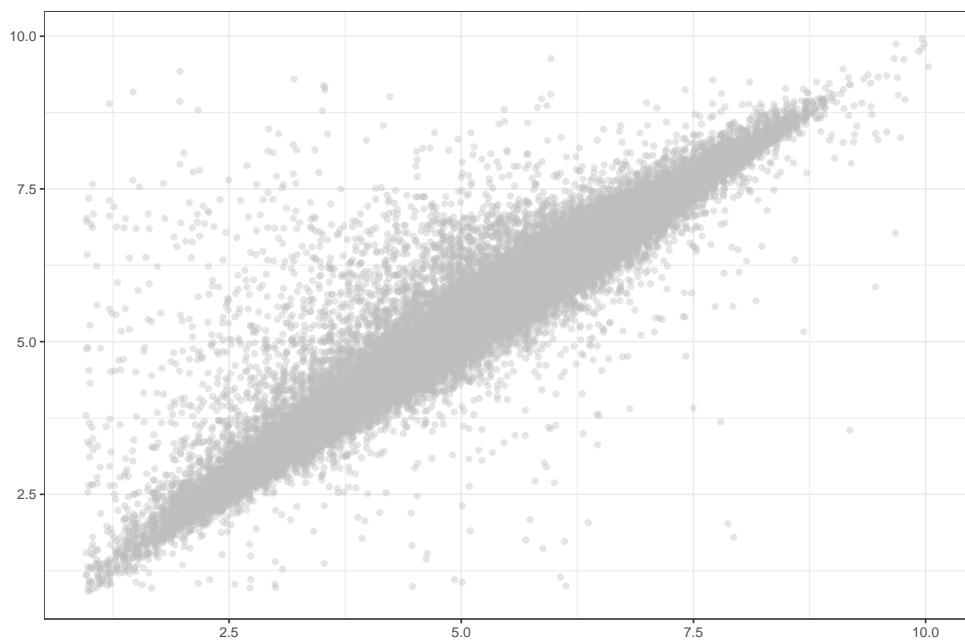


Figura 9. Classificação média de não americanos X Classificação média ponderada total.

5.2. Modelos preditivos

Após a análise exploratória dos dados, iniciamos o desenvolvimento dos modelos preditivos com o objetivo de predizer as classificações dos filmes com base nos atributos presentes no conjunto de dados. A seguir, iremos ver o resultado de cada algoritmo.

5.2.1. Naïve Bayes

O Classificador Naïve Bayes, não teve um bom desempenho, ficando com uma acurácia de 0,07 com erro absoluto médio de 2,83 e erro quadrático médio de 3,32. Na Tabela 8, são apresentados mais detalhes sobre os resultados. É interessante perceber que tanto a classe 1,0 e 10,0 possuem uma precisão de zero assim como algumas outras bem perto de zero, ou seja, isso acontece pois existem poucas ocorrências desse tipo de classe dificultando a aprendizagem do algoritmo.

Tabela 8. Resultados das métricas do classificador Naïve Bayes.

Precision	Recall	F-Measure	Classe
0	0,94	0	1,0
0,55	0,02	0,05	2,0
0,15	0,18	0,16	3,0
0,02	0,08	0,03	4,0
0,08	0,04	0,06	5,0
0,41	0,1	0,16	6,0
0,34	0,03	0,06	7,0
0,25	0,08	0,13	8,0
0,04	0,13	0,06	9,0
0	0,5	0,01	10,0

5.2.2. Random Forest

O algoritmo Random Forest, obteve um bom desempenho pois sua acurácia atingiu 0,93 com erro absoluto médio de 0,06 e erro quadrático médio de 0,07. Na Tabela 9, são apresentados mais detalhes sobre os resultados. É importante ressaltar que o algoritmo teve um bom desempenho, nas métricas apresentadas, em cada classe com exceção da classe 10,0 onde todas as métricas ficaram com zero, mais uma vez devido a pouca quantidade de amostras dessa classe.

Tabela 9. Resultados das métricas do classificador Random Forest.

Precision	Recall	F-Measure	Classe
0,92	0,75	0,83	1,0
0,92	0,94	0,93	2,0
0,89	0,91	0,90	3,0
0,91	0,91	0,91	4,0
0,91	0,92	0,91	5,0
0,95	0,95	0,95	6,0
0,95	0,95	0,95	7,0
0,96	0,93	0,94	8,0
0,85	0,72	0,78	9,0
0	0	0	10,0

Na Figura 10, temos a relação das variáveis mais preditoras, ou seja, aquelas que são mais relevantes para o algoritmo realizar a predição.

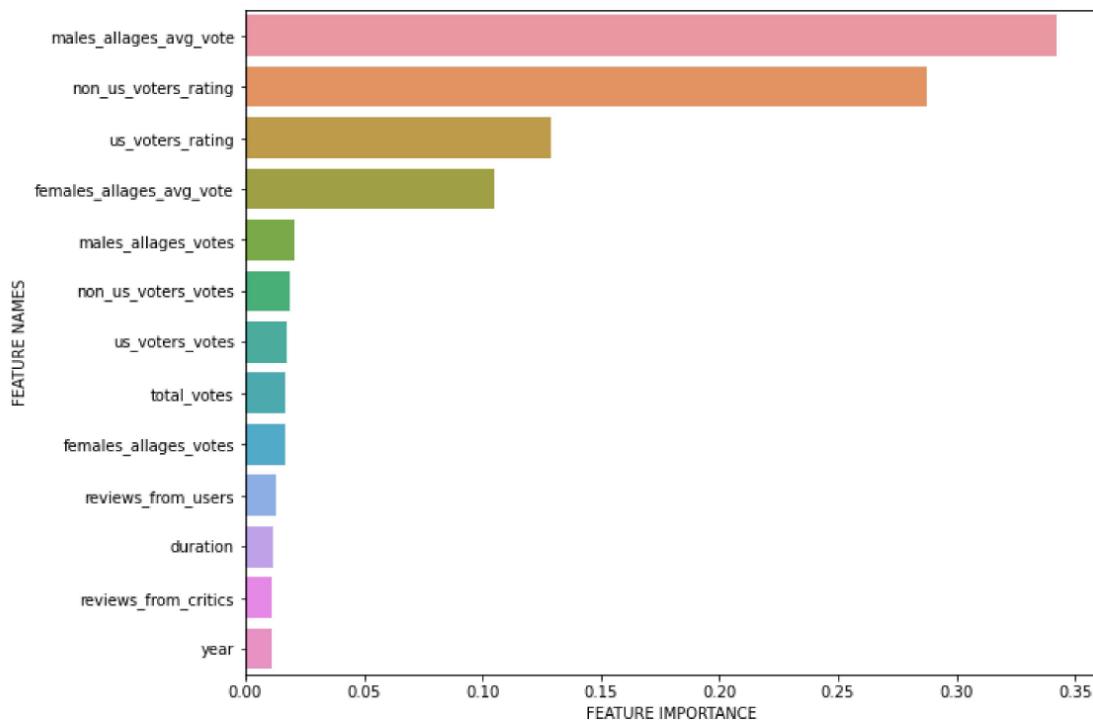


Figura 10. Variáveis mais preditoras do classificador Random Forest.

5.2.3. XGBoost

O algoritmo XGBoost, alcançou o maior desempenho comparado aos outros dois, sua acurácia atingiu 0,94 com erro absoluto médio de 0,06 e erro quadrático médio de 0,07. Na Tabela 9, são apresentados mais detalhes sobre os resultados. Nesse caso, além do bom resultado nas métricas das classe de 1,0 a 9,0 esse algoritmo também teve um bom desempenho na precisão para a classe 10,0 apesar de que seu recall foi de 0,50 mostrando que de todas as ocorrências da classe 10,0 como valor esperado apenas 50% foi classificada corretamente.

Tabela 10. Resultados das métricas do classificador XGBoost.

Precision	Recall	F-Measure	Classe
0,92	0,75	0,83	1,0
0,91	0,94	0,92	2,0
0,90	0,90	0,90	3,0
0,91	0,92	0,92	4,0
0,92	0,92	0,92	5,0
0,95	0,95	0,95	6,0
0,95	0,95	0,95	7,0
0,96	0,93	0,95	8,0
0,84	0,82	0,83	9,0
1,0	0,50	0,67	10,0

A Figura 11, elenca as variáveis mais preditoras para o algoritmo XGBoost, pode-

mos perceber que a variável classificação média de todos os usuários do sexo masculino tem o maior destaque dentre as demais.

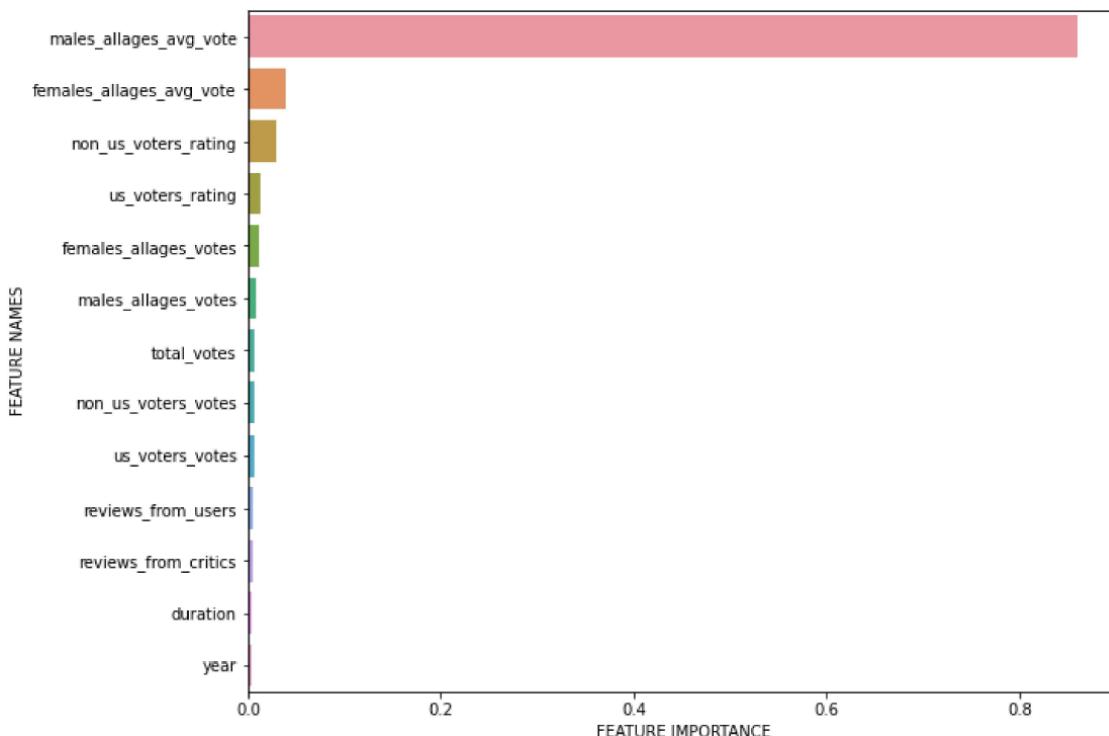


Figura 11. Variáveis mais preditoras do classificador XGBoost.

5.2.4. Resultado Geral e Efeito das Variáveis Preditoras

A Tabela 11, contém o resultado geral de cada classificador, podemos perceber claramente que o classificador XGBoost possui um melhor desempenho dentre os demais. Não apenas pelo desempenho geral, onde sua acurácia foi 0,94, não tão significante em relação ao Random Forest mas também pelas demais métricas. Nesse caso, podemos ressaltar o bom desempenho na precisão com o resultado de 0,92 e recall de 0,85 uma diferença de cerca de 19% e 6%, respectivamente, em relação ao segundo melhor classificador o Random Forest.

Tabela 11. Resultados das métricas do classificador XGBoost.

Classificador	Precision	Recall	F-Measure	MAE	MSE	Acuracy
Naïve Bayes	0,18	0,21	0,07	2,83	11,08	0,07
Random Forest	0,73	0,79	0,81	0,06	0,07	0,93
XGBoost	0,92	0,85	0,88	0,06	0,07	0,94

Além dos resultados apresentados, realizamos uma comparação do desempenho da métrica precision do melhor classificador em relação aos demais. O efeito medido foi a média da diferença entre a precisão do XGBoost em relação ao Naïve Bayes e Random Forest utilizando intervalos de confiança. A Figura 12, apresenta o resultado obtido e

podemos visualizar que a diferença entre o Random Forest e XGBoost é praticamente 0%, 95% IC[-0.02, 0.04], dessa forma não há efeito relevante na população ao escolher um dos dois modelos. Por outro lado, existe um efeito bastante relevante com cerca de 74% maior ao utilizar o XGBoost ao invés do Naïve Bayes (0,74, 95% IC[0.71, 0.77]) com uma variação pequena como visualizamos no intervalo.

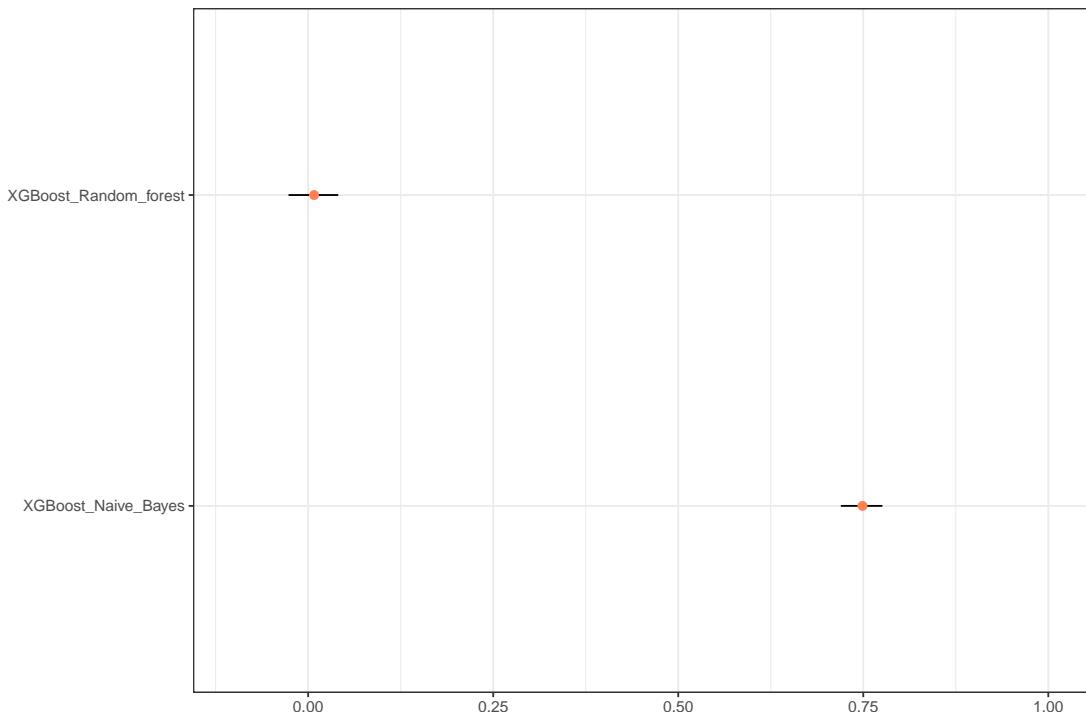


Figura 12. Diferença entre a média da precisão.

Fizemos um novo experimento, dessa vez, retiramos da base de dados as cinco primeiras variáveis apontadas como mais importantes para o algoritmo XGBoost e a sua acurácia caiu para 0,41 , ou seja, de fato as variáveis mais importantes possuem um grande peso no desempenho do classificador. Com esse resultado podemos afirmar que elas são, de fato, as variáveis mais preditoras para o modelo desenvolvido.

5.3. Conclusões

Com base nos resultados apresentados anteriormente, concluímos que o classificador Naïve Bayes não pode ser utilizado para a predição das classificações de filmes pois seu desempenho foi muito baixo. Já os algoritmos Random Forest e XGBoost obtiveram ótimos resultados mas o XGBoost conseguiu superá-lo.

Conseguimos identificar as variáveis mais preditoras para que o desempenho do classificador seja satisfatória e comprovamos que ao retirá-las da base de dados o desempenho do classificador tem uma queda considerável

6. Replicação vs. estudo original

Existem algumas diferenças entre o estudo original e a replicação:

- Base de dados utilizada na replicação foi diferente;

- Estudo original utilizou três variáveis incluindo a variável alvo, a replicação utilizou quatorze;
- Estudo original utilizou a ferramenta Weka para construção dos modelos enquanto na replicação foi utilizado R e Python;

Devido a essas diferenças principais elencadas, obtivemos um resultado diferente principalmente no classificador naïve bayes o único em comum entre os dois estudos. No estudo original esse algoritmo apresentou ótimo desempenho, atingindo F-Measure igual a 1, indicando um possível erro ou overfitting. Já em nossa replicação foi o algoritmo que apresentou o pior desempenho.

Vale ressaltar que a única contribuição em comum entre os dois estudos é apontar o algoritmo com melhor desempenho, mas, na replicação além disso, também foram elencadas quais as variáveis mais importantes para o classificador e o impacto em seu resultado quando elas não estão presentes.

7. Material para replicação

Os códigos utilizados nessa replicação encontra-se no repositório: https://github.com/jeanteixeira/fpcc2_projeto_final_doutorado

Referências

- Han, J., Kamber, M., and Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, pages 83–124.
- Jyoti, Dhawan, S., and Singh, K. (2015). Analysing user ratings for classifying online movie data using various classifiers to generate recommendations. In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, pages 295–300.
- Ricci, F., Rokach, L., and Shapira, B. (2015). Recommender systems: introduction and challenges. In *Recommender systems handbook*, pages 1–34. Springer.