

CRISP-DM

Daniel Feitosa, Felipe Alcântara, Matheus Félix, Willian Oliveira

I. ENTENDIMENTO DO NEGÓCIO

Todo documento escrito em um instante da linha do tempo incorpora aspectos linguísticos inerentes a uma determinada época. Quer seja pelo uso de signos gráficos que sejam abolidos ou mesmo por palavras ou expressões idiomáticas que caíram em desuso, o leitor do presente século que não disponha de elementos adicionais dificilmente compreenderá um texto do século passado.

A paleografia enquanto estudo da escrita antiga tem por principal propósito esclarecer o conteúdo semântico de textos antigos. Para a consecução desse fim, uma série de técnicas são empregadas e cada vez mais especializadas com possibilidades de aplicações em outros domínios. Um exemplo é a atribuição de autoria a um texto manuscrito. Essa técnica toma parte em uma análise paleográfica e tem aplicabilidade em outros contextos como análise forense e identificação de fraudes em documentos.

Existem empresas que oferecem serviços de identificação de autoria em documentos manuscritos, a grande maioria se utilizam de técnicas como análise de caligrafia e análise de assinaturas para identificar a autoria de documentos manuscritos. A análise de caligrafia é uma técnica que envolve a análise da escrita manual para identificar características individuais do escritor, como a pressão da caneta e o espaçamento entre as letras. A análise de assinaturas é uma técnica que envolve a análise da assinatura do escritor para identificar características individuais do escritor, como a forma das letras e a pressão da caneta.

Embora essas técnicas possam ser usadas para identificar a autoria de textos manuscritos, elas têm limitações e podem não ser suficientes para identificar com precisão a autoria em todos os casos. As vantagens das técnicas de aprendizagem de máquina, além da automação do processo, incluem sua capacidade de lidar com dados não lineares e sua capacidade de lidar com dados de alta dimensão. São também técnicas robustas que podem lidar com dados ruidosos. Tais características perfazem a distinção e a diferenciação do negócio proposto.

1) *Objetivos do Negócio*: Identificar de forma quantitativa e com elevado nível de certeza a autoria de conteúdos gráfico-simbólicos.

2) *Avaliação da Situação*: A identificação de autoria de forma quantitativa requer um conjunto de recursos humanos e materiais. Os recursos humanos necessários devem possuir competências técnicas em processamento e análise de dados assim como um bom entendimento de generalidades paleográficas. Os recursos materiais mínimos para empreender atividades dessa natureza são: armazenamento em disco, memória RAM e processador em quantidades compatíveis com a natureza do problema a ser resolvido.

3) *Determinação de Metas*: A determinação de metas e objetivos quantitativos na identificação da autoria de manuscritos tem por finalidade ser um parâmetro de avaliação e acompanhamento do empreendimento. Considerar o percentual de 70 por cento de identificação em relação ao total de documentos examinados parece ser plausível para o sucesso do negócio.

4) *Plano do Projeto*: Para atingir as metas e objetivos propostos podemos definir um plano em três momentos específicos: "Agora, Em seguida e Por fim".

4.1 - "Agora- Compreensão dos Dados e Engenharia de Recursos - Primeiramente, realizaremos uma análise exploratória dos dados, buscando aumentar nosso entendimento sobre as informações disponíveis para abordar o problema de negócio. Também avaliaremos as opções de pacotes e ferramentas computacionais a serem utilizadas. Durante essa fase, selecionaremos e criaremos recursos relevantes para a modelagem, aplicando técnicas de engenharia de recursos conforme as necessidades identificadas durante a análise exploratória.

4.2 - "Em seguida- Engenharia de Recursos e Modelagem - Posteriormente, planejamos desenvolver e testar modelos como K-NN, LVQ, Árvore de Decisão, SVM, Random Forest, Rede Neural MLP, Comitês de Redes Neurais Artificiais e Comitês Heterogêneos. Com base nos experimentos realizados utilizando validação cruzada, novos recursos poderão ser criados, modificados ou removidos, ajustando-se o processo de engenharia de recursos conforme necessário para melhorar o desempenho dos modelos sem comprometer sua capacidade de generalização.

4.3 - "Por fim- Avaliação e Seleção dos modelos - Após isso, realizaremos testes estatísticos para selecionar os modelos com desempenho consistentemente superior aos demais na identificação da autoria de conteúdos gráfico-simbólicos. Finalmente, aplicaremos uma predição em um conjunto de dados "Holdout" usando cada modelo, avaliando e escolhendo aqueles que conseguem generalizar as classificações para dados não utilizados durante o processo de treinamento.

II. ENTENDIMENTO DOS DADOS: ANÁLISE EXPLORATÓRIA

O conjunto de dados Avila consiste em características extraídas de 800 imagens da "Bíblia de Avila", uma cópia monumental em latim da Bíblia inteira, produzida durante o século XII entre Itália e Espanha. A análise paleográfica do manuscrito identificou a presença de 12 copistas, cuja distribuição de páginas escritas é desigual.

Cada registro do conjunto de dados é composto por 10 características, correspondendo a um grupo de 4 linhas consecutivas de texto manuscrito. Essas características, rotuladas de F1 a F10, incluem a distância entre colunas, as margens superior e inferior, a exploração, o número da linha, o ratio

modular, o espaçamento entre linhas, o peso e o número de picos. Além disso, há uma combinação do ratio modular com o espaçamento entre linhas.

Tabela I: Descrição dos atributos

ID	Name
F1	Intercolumnar distance
F2	Upper margin
F3	Lower margin
F4	Exploitation
F5	Row number
F6	Modular ratio
F7	Interlinear spacing
F8	Weight
F9	Peak number
F10	Modular ratio/interlinear spacing
Class	A, B, C, D, E, F, G, H, I, W, X, Y

Os dados já se encontram normalizados usando o método de Z-Score de normalização, o que é importante para muitos algoritmos de aprendizado de máquina que funcionam melhor ou exigem que as informações estejam em uma escala semelhante.

Os dados também já foram divididos em dois conjuntos: um conjunto de treinamento contendo 10430 amostras e um conjunto de teste contendo 10437 amostras. Não foi identificado “missing data” em nenhum dos conjuntos definidos.

A avaliação de distribuição, conforme figura 1, nos indica que a separação de dados em treinamento e teste foi feita de forma estratificada, visto que ambos datasets possuem volume similar de dados em cada classe.

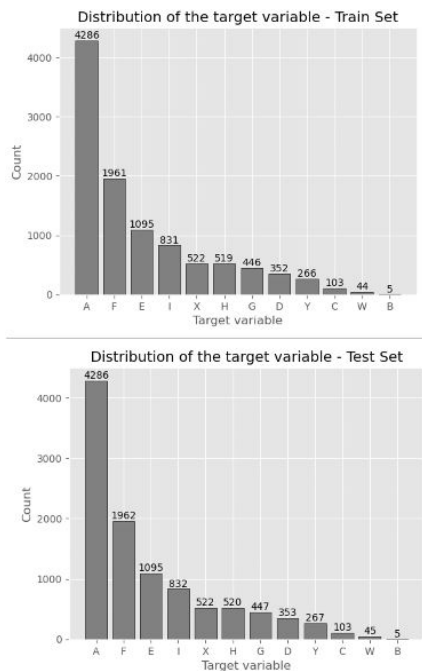


Figura 1: Distribuição dos dados

Nos nossos dados de treinamento, classe mais numerosa é a classe “A”, com 4286 observações, enquanto a classe “B” tem apenas 5 observações. Isso é uma indicação significativa de desbalanceamento de classes, que pode impactar a performance de muitos algoritmos de aprendizado de máquina.

A razão pela qual o desbalanceamento de classes pode ser problemático é porque a maioria dos algoritmos de aprendizado de máquina são projetados para maximizar a precisão geral. Eles tendem a prever preferencialmente a classe majoritária, já que isso aumenta a chance de uma previsão correta. Como resultado, a classe minoritária pode ser frequentemente mal classificada. Temos várias formas de lidar com este problema de desbalanceamento:

1-Reamostragem: Isso pode ser feito de duas maneiras - sobreamostragem da classe minoritária ou subamostragem da classe majoritária. A sobreamostragem envolve a geração de exemplos sintéticos da classe minoritária, enquanto a subamostragem envolve a remoção de exemplos da classe majoritária.

2-Adaptação de algoritmos: Alguns algoritmos permitem a atribuição de pesos diferentes às classes, de modo que a classe minoritária tem um peso maior e, portanto, um impacto maior na função de custo.

3-Uso de Métricas melhores: Em vez de usar precisão ou acurácia, outras métricas como a curva ROC AUC, F1 Score e Recall podem ser mais adequadas em casos de classes desbalanceadas.

4-Ensembles: Métodos de ensemble, como bagging e boosting, podem ser eficazes em lidar com o desbalanceamento de classes.

A análise de correlação das variáveis, apresentada na figura 2, indica que a F6 apresenta forte correlação positiva com a F10, o que é uma descoberta já esperada, visto que a F10 (Modular ratio/interlinear spacing) é derivada da divisão de F6 (Modular ratio) por F7 (Interlinear spacing). A presença de uma forte correlação entre essas duas variáveis pode ter implicações para a modelagem. Em alguns modelos, como a regressão linear, a multicolinearidade (ou seja, a alta correlação entre variáveis preditoras) pode ser problemática e interferir muito no funcionamento do modelo. Se formos utilizar modelos que captam características lineares, pode ser necessário considerar métodos para lidar com essa alta correlação, como a remoção de uma das variáveis ou a aplicação de técnicas de regularização.

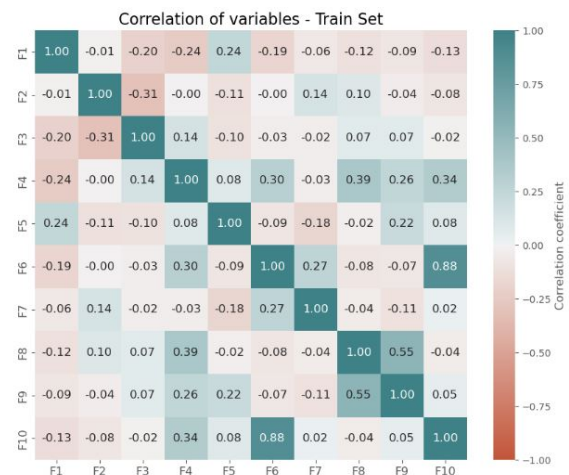


Figura 2: Correlação

III. PREPARAÇÃO DOS DADOS

O conjunto de dados foi disponibilizado de forma pré-processada, cabendo um nível de esforço mínimo nessa etapa do projeto.

As subseções a seguir detalham as decisões tomadas referentes à preparação.

A. Seleção dos Dados

Foram selecionados todos os 10 atributos numéricos (F1 à F10) e o rótulo categórico (Class) exibidos na tabela I. Não foram feitas inclusões nem exclusões de atributos nessa etapa.

B. Limpeza dos Dados

Como os dados foram disponibilizados sem “missing data” e todas as colunas de atributos numéricos haviam sido previamente normalizadas, não foi identificada a necessidade do emprego de técnicas de limpeza.

C. Construção dos Dados

No contexto de construção dos dados, foi adicionada uma transformação nos rótulos (atributo **Class**). Os caracteres designativos dos rótulos foram substituídos por valores numéricos. Esse mapeamento foi feito utilizando a classe **LabelEncoder** do pacote **sklearn.preprocessing**. Após a transformação, os novos rótulos passaram a assumir valores de 0 à 11.

Apesar do evidente desbalanceamento evidenciado nos dados de treino (ver figura 1), a inclusão de dados sintéticos ou remoção de dados existentes poderia levar a contextos indesejados. A presença simultânea de uma classe com muitas observações e outras com quase nenhuma levaria à introdução de um grande volume de dados sintéticos ou a retirada de muitos dados originais, estratégias essas indesejáveis e não incorporadas ao escopo desse trabalho.

D. Integração dos Dados

Não aplicável. O conjunto de dados foi disponibilizado de forma autocontida e já segmentado internamente em treino e teste.

E. Formatação dos Dados

Conforme pontuado na seção II, os dados foram disponibilizados com os atributos numéricos normalizados por meio da técnica de z-score. Não foram adicionados outros métodos de padronização além do supramencionado.

IV. MODELAGEM

Foram selecionados 7 modelos para variação paramétrica e obtenção dos melhores resultados no contexto do conjunto de dados selecionado. As subseções a seguir irão detalhar as configurações paramétricas escolhidas e os elementos decisórios que embasaram a construção dos modelos desenvolvidos ao longo do presente trabalho.

A. Técnica de Seleção do Modelo

Os seguintes modelos foram previamente selecionados para compor o processo decisório:

- Modelo KNN: Algoritmo utilizado em problemas de classificação, onde busca classificar cada amostra de um conjunto de dados avaliando sua distância em relação aos vizinhos mais próximos.

- Modelo LVQ: Algoritmo pertencente à família de redes neurais artificiais, utilizado para classificar e agrupar dados.

- Modelo Decision Tree: Modelo que se traduz em uma mapa dos possíveis resultados de uma série de escolhas relacionadas, permitindo a comparação de possíveis ações com base em seus custos, probabilidades e benefícios.

- Modelo SVM: Algoritmo de aprendizagem supervisionada usado para classificação ou regressão, bastante útil em problemas de classificação com muitas variáveis.

- Modelo Random Forest: Algoritmo que combina a saída de múltiplas árvores de decisão para chegar em um único resultado. Utilizado em problemas de classificação e regressão, onde é capaz de lidar com dados faltantes e variáveis categóricas.

- Modelo MLP: Modelo de Rede Neural artificial utilizado em classificação e regressão. É composto por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é composta por um conjunto de neurônios que estão conectados a neurônios da camada seguinte. O modelo lida bem com reconhecimento de fala e caracteres manuscritos bem como previsão do mercado financeiro.

- Modelo Ensemble: Combina diversos modelos de aprendizado de máquina para melhorar o desempenho preditivo. É usado em problemas de classificação e regressão e é capaz de lidar com dados ruidosos.

- Modelo XGBoost: Modelo de implementação do algoritmo Gradient Boosting utilizando árvores de decisão como base. É capaz de lidar com dados ruidosos e desbalanceados.

A partir dos modelos descritos serão construídos 2 comitês, um homogêneo e outro heterogêneo. O critério de seleção será baseado no maior valor de F1-Score encontrado.

B. Teste

Para avaliar a qualidade e validade dos modelos, serão calculadas as métricas de acurácia, precisão, recall, F1-Score e tempo de ajuste. Inicialmente os modelos serão instanciados com seus parâmetros padronizados (default) e a seguir serão executados procedimentos de tunelamento para encontrar os melhores parâmetros.

O Tuning dos Dados será empregado com o objetivo de melhorar o desempenho dos modelos por meio da otimização dos parâmetros de cada algoritmo. Será utilizado o OPTUNA tomando o F1-Score como critério de otimização com 20% dos dados para fazer o tuning e 50 iterações de busca para cada modelo.

C. Construção do Modelo

A tabela II sintetiza os resultados obtidos fazendo a variação paramétrica dos modelos KNN, GLVQ, SVC, DecisionTreeClassifier, RandomForestClassifier, XGBClassifier e

Tabela II: Variação Paramétrica

Modelo	Parâmetros	Domínio	Resultados	F1-score
KNN	n_neighbors	[1,100]	16	0.609
	weights	uniform, distance	distance	
	metric	euclidean, manhattan	manhatan	
GLVQ	activation_type	identity, sigmoid, soft+, swish	swish	0.418
	distance_type	squared_euclidean, euclidean	squared_euclidean	
	prototype_n_per_class	[1,100]	83	
SVC	c	[1e-10,1e10]	172.606797341	0.237
	kernel	poly, rbf, sigmoid	sigmoid	
	degree	[1,5]	3	
DecisionTreeClassifier	criterion	gini, entropy	entropy	0.728
	splitter	best, random	best	
	max_depth	[1,100]	16	
	min_samples_leaf	[1,100]	6	
	min_samples_split	[2,100]	17	
RandomForestClassifier	n_estimators	[1,25]	15	0.594
	criterion	gini, entropy	gini	
	max_depth	[1,100]	20	
	min_samples_leaf	[1,100]	5	
	min_samples_split	[2,100]	34	
XGBClassifier	max_depth	[1,100]	31	0.782
	learning_rate	[1e-10,1e10]	8.9834e-09	
	n_estimators	[1,25]	18	
	gamma	[1e-10,1e10]	0.06354	
MLPClassifier	activation	identity, logistic, tanh, relu	tanh	0.702
	solver	sgd, adam	sgd	
	n_layers	[1,5]	1	
	hidder_layers_sizes	[1,100]	72	
	learning_rate_type	constant, invscaling, adaptative	adaptative	
	learning_rate_init	[1e-10,1e10]	0.7514704	

MLPClassifier. A coluna **Parâmetros** exibe o conjunto de parâmetros escolhidos para variação e a coluna **Domínio** sintetiza o domínio de interesse assumido por cada parâmetro. Os melhores resultados para cada domínio são mostrados na coluna **Resultados** e o F1-score obtido para cada modelo é exibido na coluna **F1-score**.

Foram construídos dois modelos de comitê com 3 classificadores para identificar a autoria de conteúdos gráfico-simbólicos: um comitê heterogêneo e outro de Redes Neurais Artificiais. O primeiro foi construído a partir dos 3 classificadores com melhores parâmetros e maior **F1-score** (Decision-Tree, XGBoost e RandomForestClassifier) dos 7 analisados. O segundo, a partir de 3 MLPs com mesma configuração de melhores parâmetros encontrados na tabela II.

D. Avaliação dos Modelos

O F1-score dos modelos de comitê constuídos com base na estratégia descrita em IV-C são mostrados na tabea III. Nela é possível evidenciar o melhor resultado obtido pelo comitê heterogêneo sem tuning.

Tabela III: F1-Score dos modelos comitês

Comitê	F1-Score	
	Sem Tuning	Com Tuning
Heterogêneo	0.974074	0.891990
RNAs	0.8090008	0.949329

V. AVALIAÇÃO

A seguir serão apresentados os principais resultados para o conjunto de modelos descrito na seção IV. Como verificado no notebook 03. Modeling.ipynb, disponibilizado em conjunto

com este relatório, a busca paramétrica executada pelo op-tuna com 50 testes não apresentou melhores resultados em comparação as configurações padronizadas dos modelos, toda a análise que se segue será baseada no cenário de modelos base.

A. Avaliação dos Resultados

A comparação entre os 7 classificadores e os comitês gerados a partir deles foi feita a partir das métricas de acurácia, precisão, recall, f1-score e tempo de treino (fit_time). A tabela IV exibe os resultados encontrados (os melhores resultados são exibidos em **negrito**).

Tabela IV: Métricas dos modelos

Model	Métricas - Teste			
	F1	Precision	Accuracy	Recall
Decision Tree	0.931462	0.932739	0.961486	0.933504
Comitê Heterogêneo	0.983268	0.986929	0.992128	0.981561
Comitê Homogêneo	0.772935	0.825640	0.779137	0.746554
GLVQ	0.371382	0.409923	0.544813	0.434515
KNN	0.671850	0.757833	0.721572	0.624547
MLP	0.763841	0.813082	0.772943	0.738424
Random Forest	0.969104	0.979193	0.979990	0.960124
SVC	0.560026	0.754558	0.670853	0.521846
XGBoost	0.987463	0.987404	0.996750	0.989765

Da análise da tabela IV, é possível observar que o modelo com o melhor desempenho geral é o XGBoost. Este apresenta o maior F1 score (0.987), precisão (0.987), acurácia (0.997) e recall (0.990) entre todos os modelos. Apesar do alto desempenho, o tempo de treinamento deste modelo é mediano quando comparado aos demais.

A Árvore de Decisão (Decision Tree) merece destaque por ter o segundo maior F1 score, precisão e recall, ficando

atrás apenas do XGBoost. A vantagem adicional deste modelo está no menor tempo de treinamento entre todos os modelos, garantindo eficiência na aplicação.

O modelo com o pior desempenho em todas as métricas de avaliação é o GLVQ, que também possui um dos maiores tempos de treinamento, indicando baixa eficiência.

O KNN tem o menor tempo de treinamento, porém seu desempenho é apenas mediano quando comparado aos outros modelos. Este pode ser uma boa escolha quando o tempo de treinamento é uma restrição importante.

Os modelos SVC e MLP apresentam desempenho divergente. O SVC, apesar de ter um tempo de treinamento inferior, apresenta um desempenho inferior ao MLP em todas as métricas de avaliação. O SVC possui uma precisão relativamente alta em comparação ao seu F1 score e recall, o que indica que o modelo possui menos falsos positivos, mas mais falsos negativos.

Os modelos Ensemble apresentam desempenhos divergentes. O Comitê Heterogêneo, composto por uma Árvore de Decisão, uma Random Forest e um XGBoost, alcança um desempenho levemente inferior ao XGBoost, mostrando-se um bom modelo. Já o Comitê Homogêneo de MLPs apresenta um desempenho significativamente inferior aos demais modelos, com F1 score de 0.773, precisão de 0.826 e acurácia de 0.779. Além disso, o tempo de treinamento deste último modelo é o maior entre todos os modelos.

Por fim, a Floresta Aleatória (Random Forest) se destaca por um bom desempenho geral, com F1 score de 0.969, precisão de 0.979, acurácia de 0.980 e recall de 0.960. Além disso, seu tempo de treinamento é relativamente baixo em comparação com os outros modelos, tornando-o um modelo eficiente e de bom desempenho.

A figura 3 mostra os primeiros pontos de quebra utilizados na construção da árvore.

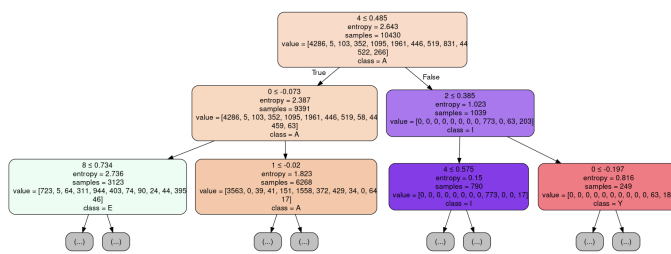


Figura 3: Árvore de Decisão - primeiros níveis

Da análise da árvore, é possível perceber que as classes com maior ganho de informação são: A, E e Y. Também é possível identificar que os principais atributos para o processo discriminativo da árvore de decisão foram os atributos F1, F2, F3, F5 e F9, sendo estes a Distância Intercolunar, a Margem superior, a Margem inferior, o Número da linha e o Número de picos nas letras, respectivamente.

Através da inspeção dos gráficos de box-plot, percebemos que os modelos Random Forest, XGBoost e o Comitê Heterogêneo se destacam dos demais, tanto pela sua performance superior, evidenciada pelas médias mais elevadas nas métricas avaliadas, quanto pela sua robustez, evidenciada pela menor

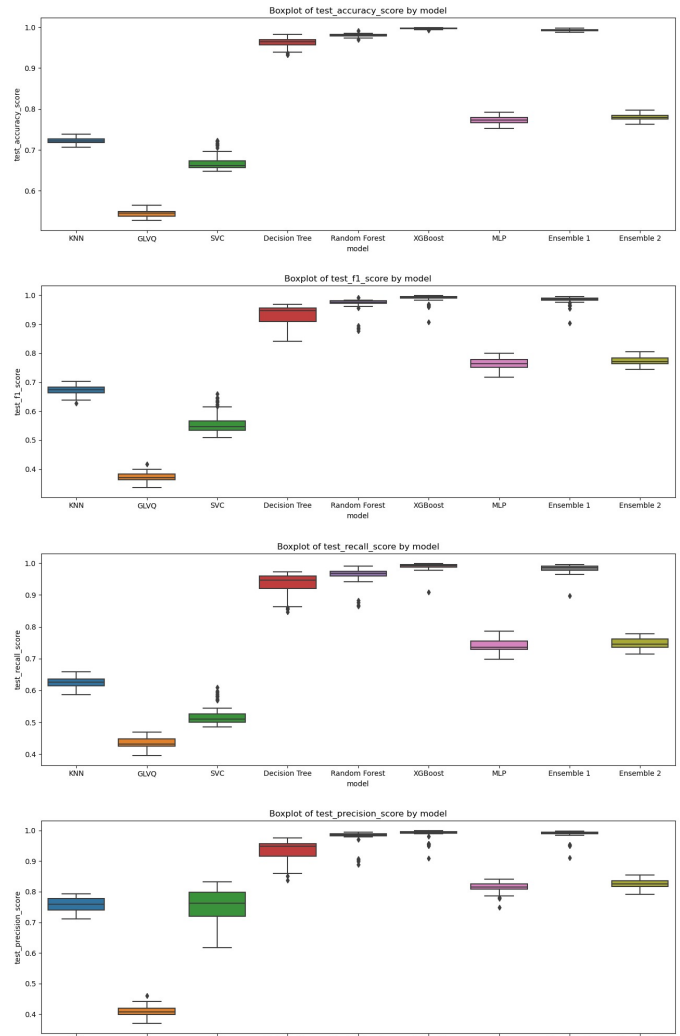


Figura 4: Boxplots

variação dessas métricas durante a execução de validação cruzada com 10 folds e 5 splits.

Os resultados sugerem que os modelos Random Forest, XGBoost e o Comitê Heterogêneo são mais estáveis e consistentes, dado que tendem a manter uma performance similar mesmo quando treinados com diferentes sub-amostras do nosso conjunto de dados.

Entretanto, é fundamental sublinhar que essas observações preliminares necessitam de confirmação através dos testes estatísticos não paramétricos de Friedman e o Post-Hoc Nemenyi. Esses testes nos ajudarão a determinar se as diferenças de desempenho entre os modelos são estatisticamente significativas. Dessa forma, estaremos aptos a fazer afirmações com 95% de confiança, ou seja, admitindo uma taxa de erro de 5%.

O teste de Friedman, exposto na tabela V, indica que há uma diferença estatisticamente significativa entre os grupos, de forma que utilizamos o teste post-hoc de Nemenyi para descobrir exatamente onde essas diferenças ocorrem.

A hipótese nula, neste caso, é que não há diferença significativa entre os grupos. Se o p-valor for inferior a um nível

Tabela V: Friedman Teste

Score	Friedman Chi Square	p-value
test_f1_score	77.760	0.000
test_precision_score	76.053	0.000
test_accuracy_score	79.333	0.000
test_recall_score	79.120	0.000

de significância de 0,05, então rejeitamos a hipótese nula e concluímos que há uma diferença estatisticamente significativa entre os grupos. Se o p-valor for superior a esse nível, não podemos rejeitar a hipótese nula e concluímos que não há diferença estatisticamente significativa.

Ao analisar os resultados presentes na tabela VI, podemos observar os seguintes resultados:

- KNN (Modelo 0) tem diferença estatisticamente significativa em comparação com RandomForest, XGBoost e Ensemble 1.
- GLVQ (Modelo 1) tem diferença estatisticamente significativa em comparação com DecisionTree, RandomForest, XGBoost, Ensemble 1 e Ensemble 2.
- SVC (Modelo 2) tem diferença estatisticamente significativa em comparação com DecisionTree, RandomForest, XGBoost e Ensemble 1.
- DecisionTree (Modelo 3) tem diferença estatisticamente significativa em comparação com GLVQ, SVC, RandomForest, XGBoost, Ensemble 1 e Ensemble 2.
- RandomForest (Modelo 4) tem diferença estatisticamente significativa em comparação com todos os modelos, exceto XGBoost e Ensemble 1.
- XGBoost (Modelo 5) tem diferença estatisticamente significativa em comparação com todos os modelos, exceto RandomForest e Ensemble 1.
- MLP (Modelo 6) tem diferença estatisticamente significativa em comparação com XGBoost e Ensemble 1.
- Comitê Heterogêneo (Modelo 7) tem diferença estatisticamente significativa em comparação com todos os modelos, exceto RandomForest e XGBoost.
- Comitê Homogêneo (Modelo 8) tem diferença estatisticamente significativa em comparação com GLVQ, XGBoost e Ensemble 1.

Com base nesses testes e nas análises dos box-plots e métricas médias, os modelos Random-Forest e XGBoost apresentaram resultados estatisticamente superiores aos demais, e o Comitê Heterogêneo, composto por modelos de Decision Tree, Random Forest e XGBoost, também conseguiu performar de forma estatisticamente similar aos melhores classificadores.

A partir da análise das matrizes de confusão dos modelos treinados, exposto na figura 5, observamos uma tendência distinta entre os modelos baseados e não baseados em árvores. Modelos não baseados em árvores parecem ser mais afetados pelo desbalanceamento do conjunto de dados. Tais modelos tendem a classificar um maior número de observações como Classe A, a classe predominante. Por outro lado, os modelos baseados em árvore demonstram uma maior eficácia na classificação de classes menos representadas no conjunto de dados. Isso ilustra sua capacidade superior para lidar com a disparidade de representação entre as classes.

VI. CONCLUSÃO

Identificar a autoria de documentos manuscritos é uma tarefa tediosa e demanda um tempo considerável se for feita de forma manual. A automação proveniente do uso de IA pode trazer uma precisão significativa aliada à redução do custo homem-hora empregado para essa finalidade.

Da análise do emprego de IA no contexto do negócio, métodos baseados em árvores tendem a ter melhor resultado e testes realizados apontam que é possível obter uma acurácia de classificação superior a 97% em alguns modelos.

REFERÊNCIAS

- [1] Ackel, Antonio. (2021). Abordagens digitais para estudos de paleografia: desafios, atualidade, desdobramentos. LaborHistórico. 7. 100-120. 10.24206/lh.v7i3.42771.
- [2] Melhores Bibliotecas Python para Aprendizagem de Máquinas - Artigo disponível em <https://datascience.eu/pt/programacao/melhores-bibliotecas-python-para-aprendizagem-de-maquinas/>
- [3] As 30 melhores bibliotecas e pacotes Python para Iniciantes - Artigo disponível em <https://terminalroot.com.br/2019/12/as-30-melhores-bibliotecas-e-pacotes-python-para-iniciantes.html>

Tabela VI: Nemenyi Teste

	0	1	2	3	4	5	6	7	8
0	1.000000	0.759750	0.900000	0.216940	0.017589	0.001000	0.900000	0.001000	0.809296
1	0.759750	1.000000	0.900000	0.001042	0.001000	0.001000	0.216940	0.001000	0.038985
2	0.900000	0.900000	1.000000	0.023112	0.001000	0.001000	0.710204	0.001000	0.301968
3	0.216940	0.001042	0.023112	1.000000	0.900000	0.611110	0.759750	0.660658	0.900000
4	0.017589	0.001000	0.001000	0.900000	1.000000	0.900000	0.216940	0.900000	0.611110
5	0.001000	0.001000	0.001000	0.611110	0.900000	1.000000	0.013263	0.900000	0.099098
6	0.900000	0.216940	0.710204	0.759750	0.216940	0.013263	1.000000	0.017589	0.900000
7	0.001000	0.001000	0.001000	0.660658	0.900000	0.900000	0.017589	1.000000	0.122859
8	0.809296	0.038985	0.301968	0.900000	0.611110	0.099098	0.900000	0.122859	1.000000

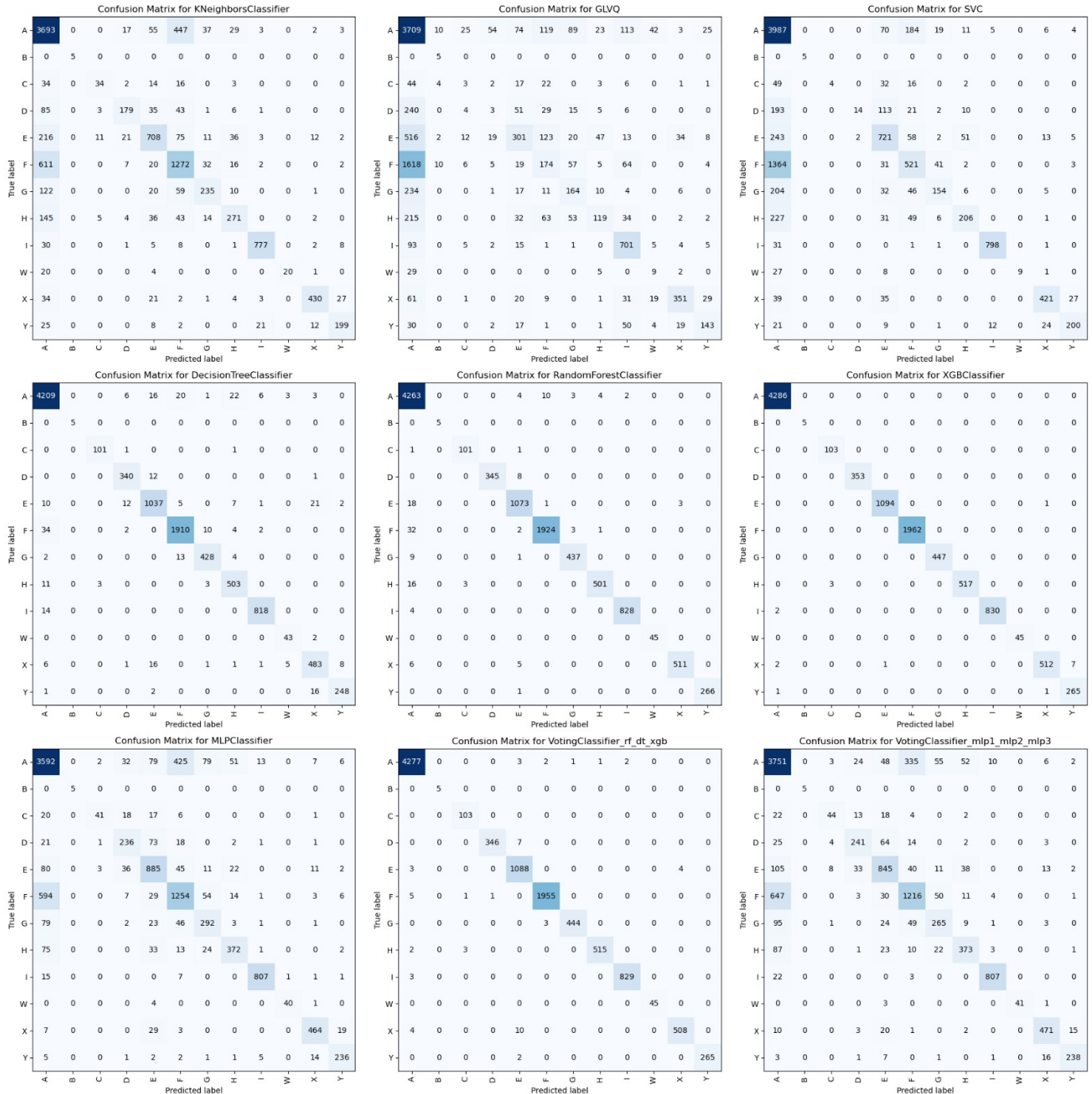


Figura 5: Matriz de Confusão