

COMPONENTE CURRICULAR:	Projeto Aplicado II
GRUPO	Grupo 13
ALUNOS	
Thaís Cristine de Andrade Gomes	10721642
Paulo Ricardo de Oliveira Ramos	10721464
Lucas Iglezias dos Anjos	10433522
Oscar Augusto de Oliveira Luz	10435099

Apresentação de Produtos e Storytelling



Sumário

1. Introdução	3
2. Aplicação do Método Analítico na Base de Dados	3
3. Medidas de Acurácia	3
4. Produto Gerado	4
5. Modelo de Negócio Proposto	4
6. Esboço do Storytelling	5
7. Repositório do Projeto	6

Introdução

Esta terceira etapa do Projeto Aplicado II tem como foco a aplicação prática do modelo analítico desenvolvido na fase anterior, com o objetivo de consolidar resultados, apresentar um produto funcional e propor um modelo de negócio viável. A base de dados utilizada contém mais de 670 mil avaliações do aplicativo Tinder, e o método escolhido foi o classificador Multinomial Naive Bayes.

Aplicação do Método Analítico na Base de Dados

O método analítico definido na etapa anterior foi o classificador supervisionado Multinomial Naive Bayes, aplicado sobre a base de dados textual composta por mais de 670 mil avaliações de usuários do aplicativo Tinder, extraídas da Google Play Store. A base foi previamente tratada e vetorizada, utilizando o modelo de representação textual TF-IDF, conforme as etapas já executadas e documentadas anteriormente.

Para aplicação prática do modelo, utilizamos os dados vetorizados de treino (`X_train`, `y_train`) e de teste (`X_test`, `y_test`), todos salvos no formato `.pkl`, possibilitando fácil reaproveitamento e escalabilidade do pipeline.

O modelo foi treinado com os dados categorizados em cinco classes de sentimento: **Muito Positiva**, **Positiva**, **Neutra**, **Negativa** e **Muito Negativa**. A escolha do Multinomial Naive Bayes se deu pela sua simplicidade, eficiência em textos vetorizados e resultados consistentes em problemas de classificação multiclasse.

Medidas de Acurácia

O desempenho do modelo foi avaliado com base em métricas tradicionais de classificação: acurácia, F1-score por classe e matriz de confusão. Os principais resultados foram:

- **Acurácia geral:** 67,76%
- **F1-score ponderado:** 0,59
- **F1-score médio geral:** 0,34

Esses números demonstram que o modelo teve bom desempenho nas classes mais extremas (Muito Positiva e Muito Negativa), que apresentaram maiores volumes de dados e

características mais marcantes. Já as classes intermediárias (Neutra, Negativa, Positiva) tiveram desempenho inferior, indicando que o modelo tem mais dificuldade para diferenciar sentimentos ambíguos ou menos expressivos.

A matriz de confusão confirmou essa tendência, mostrando acertos expressivos nos extremos e maior confusão entre as classes centrais.

Esse comportamento é coerente com as limitações do Naive Bayes, que assume independência entre atributos e pode ser afetado pela subjetividade presente em textos neutros.

Produto Gerado

O produto desenvolvido até esta etapa é um pipeline de classificação de sentimentos textual completo e funcional, capaz de processar e classificar grandes volumes de comentários de usuários. O pipeline inclui:

- I. **Pré-processamento de texto:** remoção de stopwords, limpeza de caracteres especiais, normalização de caixa e remoção de duplicatas.
- II. **Categorização de sentimentos:** usando mapeamento com base em polaridade.
- III. **Vetorização com TF-IDF:** modelo salvo em `tfidf_vectorizer.pkl`.
- IV. **Classificação com Naive Bayes:** modelo salvo em `multinomial_nb_model.pkl`.
- V. **Separação e salvamento dos conjuntos de treino e teste:** `X_train`, `X_test`, `y_train`, `y_test`.

Todos os scripts foram organizados em arquivos `.py` e `.ipynb`, com lógica modular para facilitar manutenção e reuso. A base de dados categorizada (`tinder_google_play_categorizado.csv`) e limpa (`tinder_google_play_limpo.csv`) também foi versionada no repositório, garantindo rastreabilidade.

Esse produto já é capaz de ser reutilizado em outros projetos com dados similares, bastando substituir a base de entrada por novos comentários em português.

Modelo de Negócio Proposto

A proposta do modelo de negócio é transformar o classificador de sentimentos em um produto de fácil acesso para empresas que desejam entender melhor seus usuários ou

colaboradores a partir de textos espontâneos, como comentários em redes sociais, avaliações de produtos ou respostas de pesquisas internas.

O produto seria uma ferramenta integrada ao Excel, onde o usuário seleciona uma coluna com comentários e a IA preenche automaticamente outra coluna com o sentimento detectado (Muito Negativo a Muito Positivo).

- **Principais segmentos que podem se beneficiar:**
 - **SAC (Serviço de Atendimento ao Cliente):** análise de reclamações e elogios para ajustar processos e atendimento.
 - **Marketing:** avaliação do impacto emocional de campanhas, produtos e posicionamentos.
 - **Recursos Humanos:** mapeamento de satisfação de colaboradores em pesquisas qualitativas.
 - **Gerentes de Produto:** identificação de falhas e oportunidades em novas funcionalidades ou lançamentos.

- **Modelo de receita:**
 - **Plano Gratuito:** até 500 análises por mês.
 - **Plano Profissional:** até 10.000 análises por mês, com suporte básico.
 - **Plano Corporativo:** ilimitado, com relatórios automáticos e personalização por setor.

A ferramenta pode ser distribuída por meio de um site simples com tutoriais, planos de uso e interface de upload de planilhas.

Esboço do Storytelling

Nossa narrativa parte da seguinte provocação: o que os usuários realmente pensam sobre os serviços que usam? Em um cenário onde dados são gerados a cada segundo, as empresas ainda têm dificuldade em compreender o sentimento por trás das palavras.

Analisando mais de 670 mil comentários sobre o Tinder, construímos uma solução capaz de transformar texto livre em informação organizada. Não apenas classificamos sentimentos, mas demos contexto a eles, revelando padrões, tensões e oportunidades invisíveis a olho nu.



Com um modelo robusto e um pipeline reutilizável, criamos um produto que pode ser facilmente adaptado para qualquer negócio que lide com clientes e precise ouvir o que eles têm a dizer.

Nosso storytelling, portanto, será estruturado da seguinte forma:

1. **Problema:** empresas não conseguem lidar com grandes volumes de feedback textual.
2. **Solução:** pipeline de classificação de sentimentos com Naive Bayes.
3. **Resultados:** métricas sólidas e produto funcional.
4. **Aplicação prática:** Excel plugin para SAC, marketing e RH.
5. **Impacto futuro:** democratizar o uso de IA para escuta ativa e tomada de decisão.

Repositório do Projeto

Todos os arquivos do projeto, incluindo scripts Python, notebooks, bases de dados originais e tratadas, vetores vetorizados e modelos treinados estão disponíveis de forma organizada no seguinte link: <https://github.com/lcsigzs/Projeto-Aplicado-II>.