

Tema:

**Multi-sentiment rule-based model for sentiment analysis of movie reviews**

## Introdução

Entender os sentimentos, as opiniões, das outras pessoas é essencial para ter boas interações com humanos. Quando duas pessoas se falam, elas analisam as expressões do rosto, as emoções, a entonação da voz do interlocutor, além do contexto da discussão e do sentido das frases, para entender as opiniões e os sentimentos expressos. Empatia tem grande importância neste processamento.

Com o aumento do processo de digitalização do mundo, humanos têm mais e mais interações com sistemas digitais. Contudo, a maioria das interações que se tem com sistemas computacional não levam em consideração os sentimentos expressos pelos humanos.

## Proposta

A proposta deste projeto é de avaliar diferentes modelos baseados na teoria já existente sobre a análise de sentimentos de textos. O objetivo é desenvolver um algoritmo para prever a avaliação quantitativa expressa por uma pessoa sobre um filme com base na opinião dada no texto de crítica que ela escreveu sobre o filme.

## Metodologia

O projeto foi programado com a linguagem Python, desenvolvido a partir das bibliotecas de Machine Learning TensorFlow, Keras e Scikit-Learn. Para gerar features a partir dos textos, foram usados léxicos de análise de sentimentos (VADER e NRC). As bases de dados da IMDB e do site Rotten Tomatoes foram usadas para treinar nossos modelos avaliados.

O processo de desenvolvimento foi iterativo. O objetivo era de experimentar várias arquiteturas de modelos e de avaliá-los para determinar a arquitetura com melhor desempenho.

Primeiro, nós foram analisar as críticas textuais de filmes, fornecidos por pessoas, nas duas bases de dados. A seguir, criou-se iterativamente modelos.

**Integrantes:** Donatien Ferdinand Marie Joseph Dujoncquoy

**Professor Orientador:** Anna Helena Reali Costa  
**Co-orientador:** Francisco Caio Lima Paiva

## Arquitetura

As crítica escrita pelo usuário é processada e codificada primeiro, antes de extrair métricas de sentimentos e de transformá-la ela em representação vetorial. As métricas e a representação vetorial são processadas pelo modelo de classificação para prever a nota da crítica. A figura 1, mostra a esquema de arquitetura proposta.

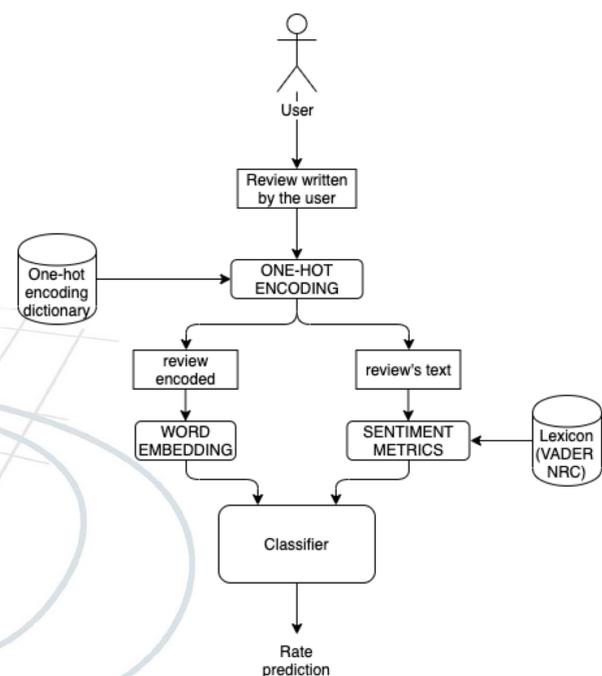


Figura 1

## Resultados

Os modelos, criados para fazer uma classificação binária das críticas (positiva ou negativa) têm bons resultados, mais de 85% de precisão (figura 2). Quando mais classes são classificadas, menor a precisão do classificador.

O impacto dos léxicos na análise de sentimentos foram menores que o esperado e as representações vetoriais do texto das críticas têm um impacto muito grande na precisão dos modelos.

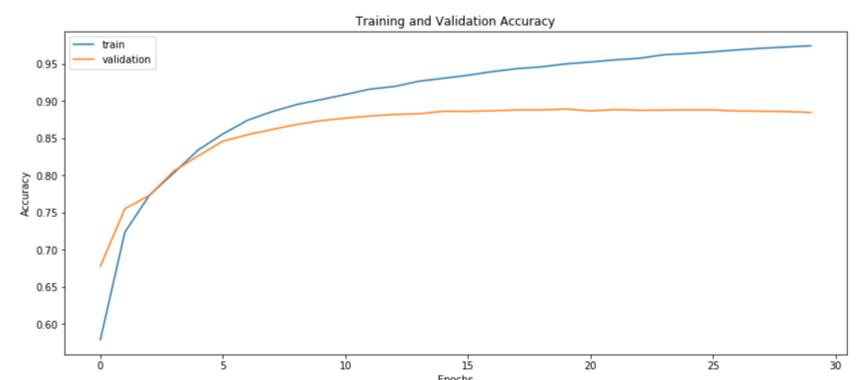


Figura 2 – Precisão do modelo binário durante treino e validação