

Relatório Crítico: Análise de Sentimentos em Tom Sawyer

Gustavo Dutra

29 de setembro de 2025

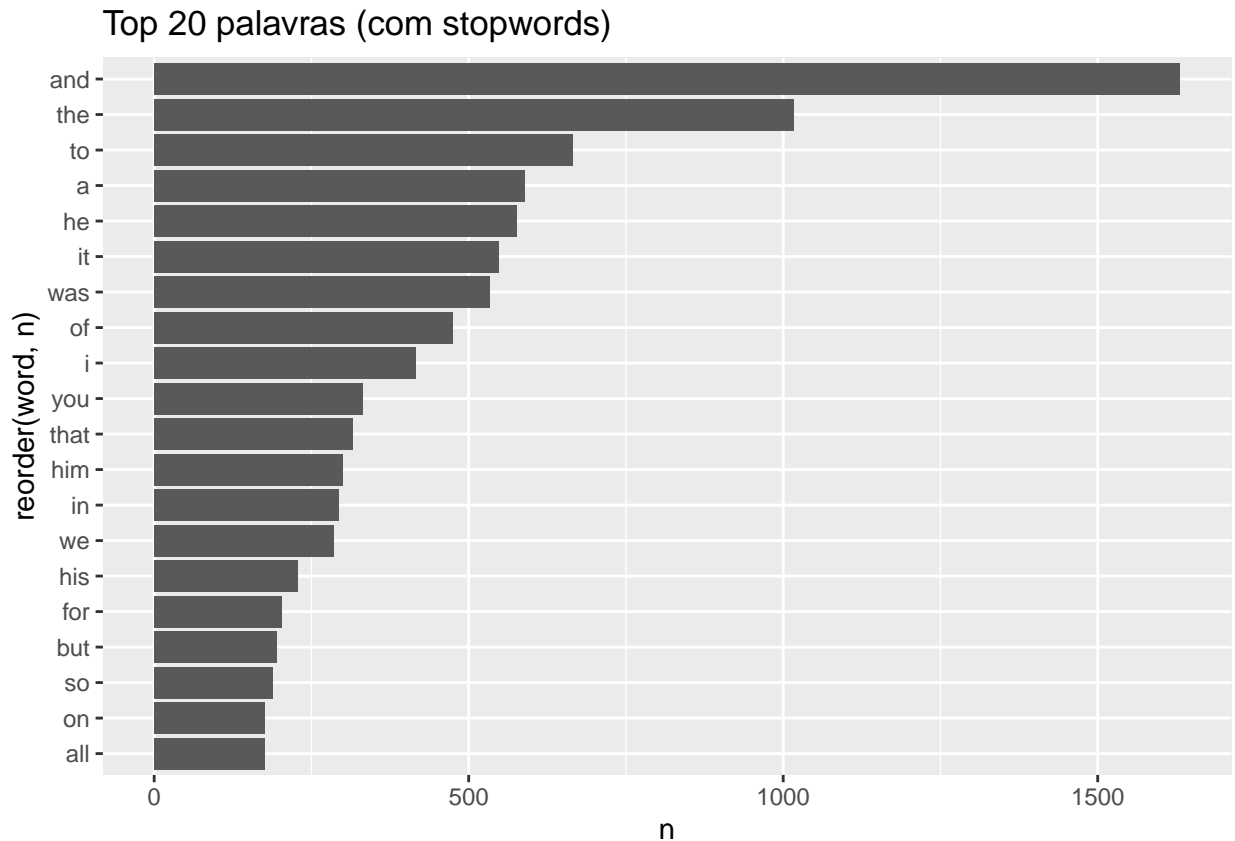
Relatório Crítico

Este relatório apresenta uma análise crítica dos resultados obtidos na atividade de mineração de texto e análise de sentimentos aplicada ao romance “The Adventures of Tom Sawyer”, de Mark Twain.

Análise dos Gráficos Gerados

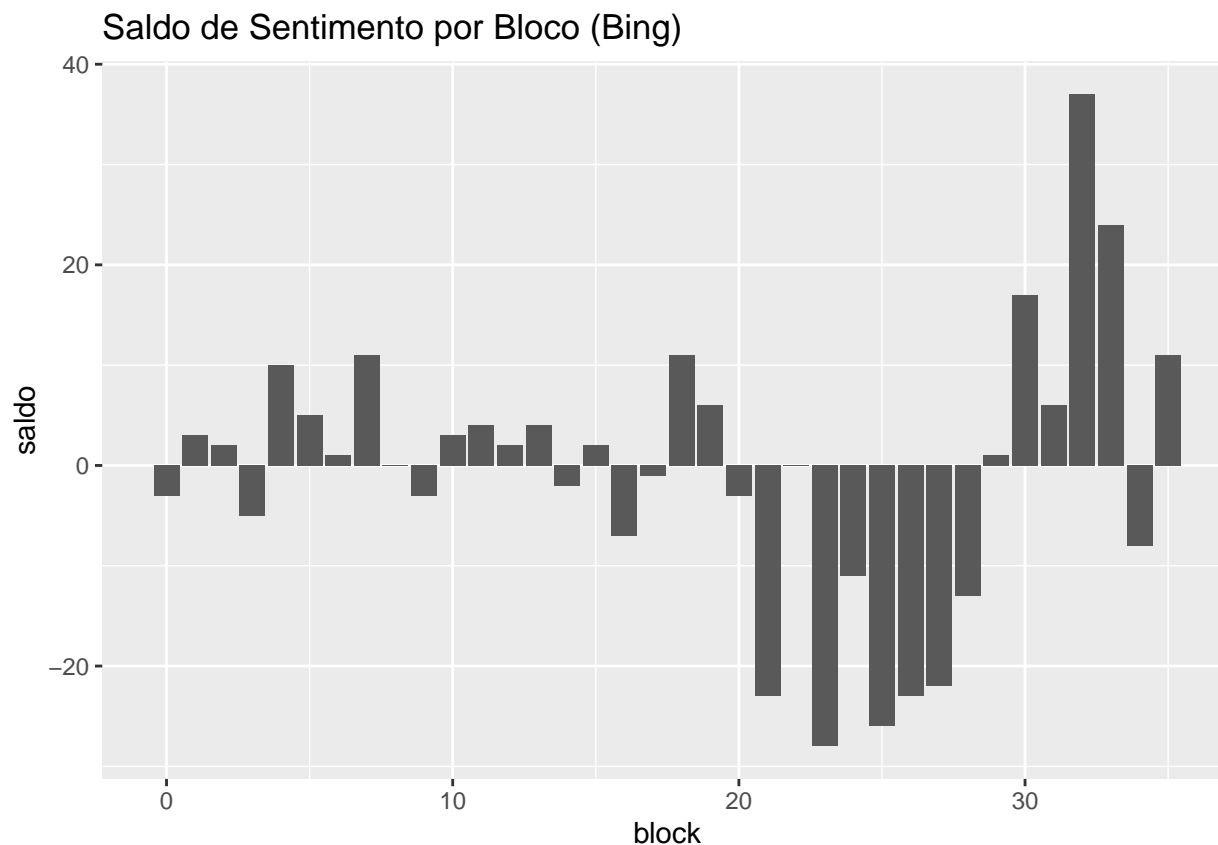
1. Top 20 Palavras (com e sem stopwords)

Os gráficos de frequência revelam padrões importantes sobre o vocabulário do romance. Com stopwords, observamos que palavras funcionais como “the”, “and”, “to” dominam, o que é esperado em textos em inglês. Isso reforça a necessidade de removê-las para análises semânticas. Sem stopwords, os nomes dos personagens principais, como “tom”, “huck” e “jim”, aparecem entre as mais frequentes, mostrando o foco narrativo. Palavras como “river”, “boy” e “uncle” indicam temas recorrentes e ambientação. É interessante notar que, mesmo após a remoção de stopwords, ainda há termos que podem ser considerados ruído, como palavras de contexto ou narração, o que sugere que uma etapa adicional de filtragem (como lematização ou análise de entidades) pode refinar ainda mais a análise.



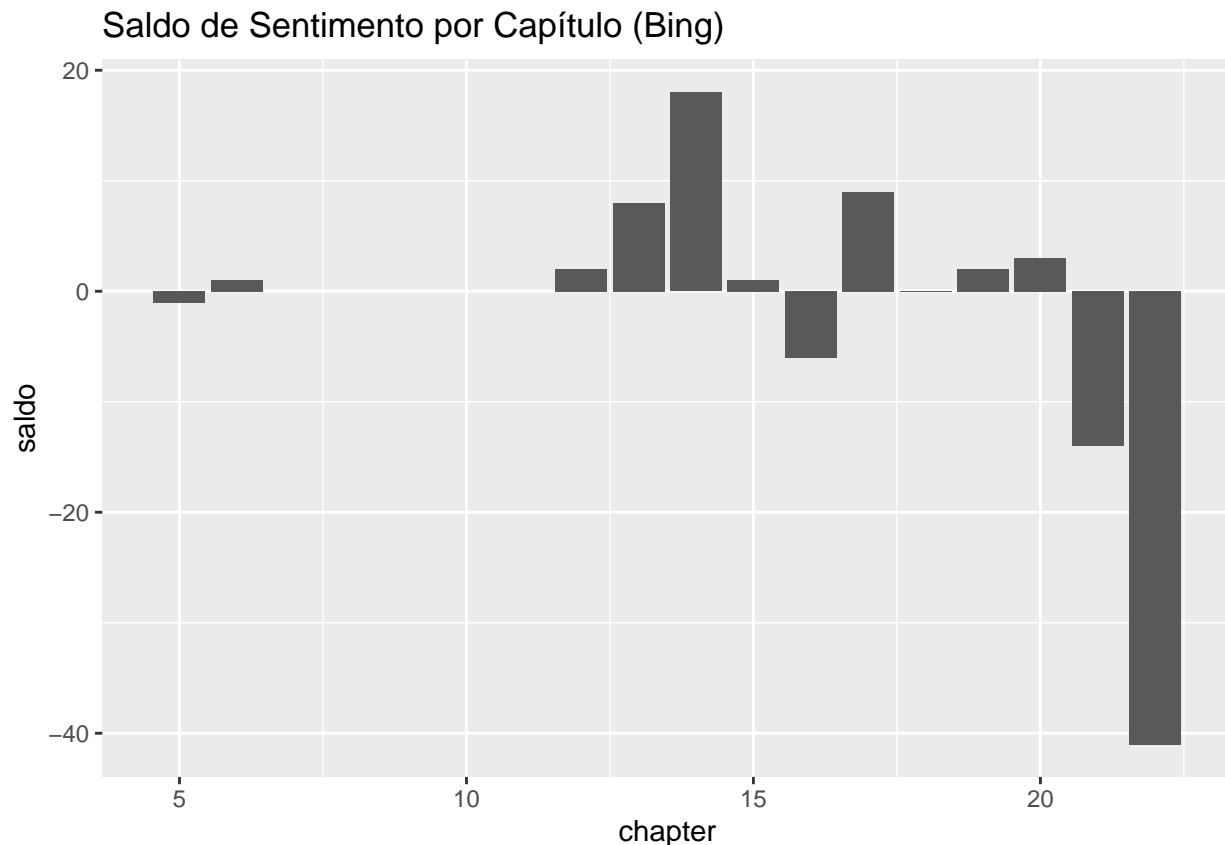
2. Saldo de Sentimento por Bloco (Bing)

O gráfico de barras por blocos de 80 linhas permite observar a dinâmica emocional do texto em alta resolução. Blocos com saldo positivo geralmente correspondem a trechos de aventura, brincadeiras ou interações amigáveis entre os personagens. Já os blocos negativos tendem a coincidir com conflitos, punições ou situações de perigo. Essa granularidade é útil para identificar rapidamente onde ocorrem mudanças bruscas de tom, como a transição de um capítulo leve para um momento de tensão. Por exemplo, é possível visualizar picos negativos em blocos que abordam o julgamento de Muff Potter ou o desaparecimento de Tom e Becky, enquanto picos positivos aparecem em passagens de celebração ou descoberta. Essa análise mostra que o romance alterna entre episódios de leveza e drama, refletindo a estrutura clássica de aventuras infantojuvenis.



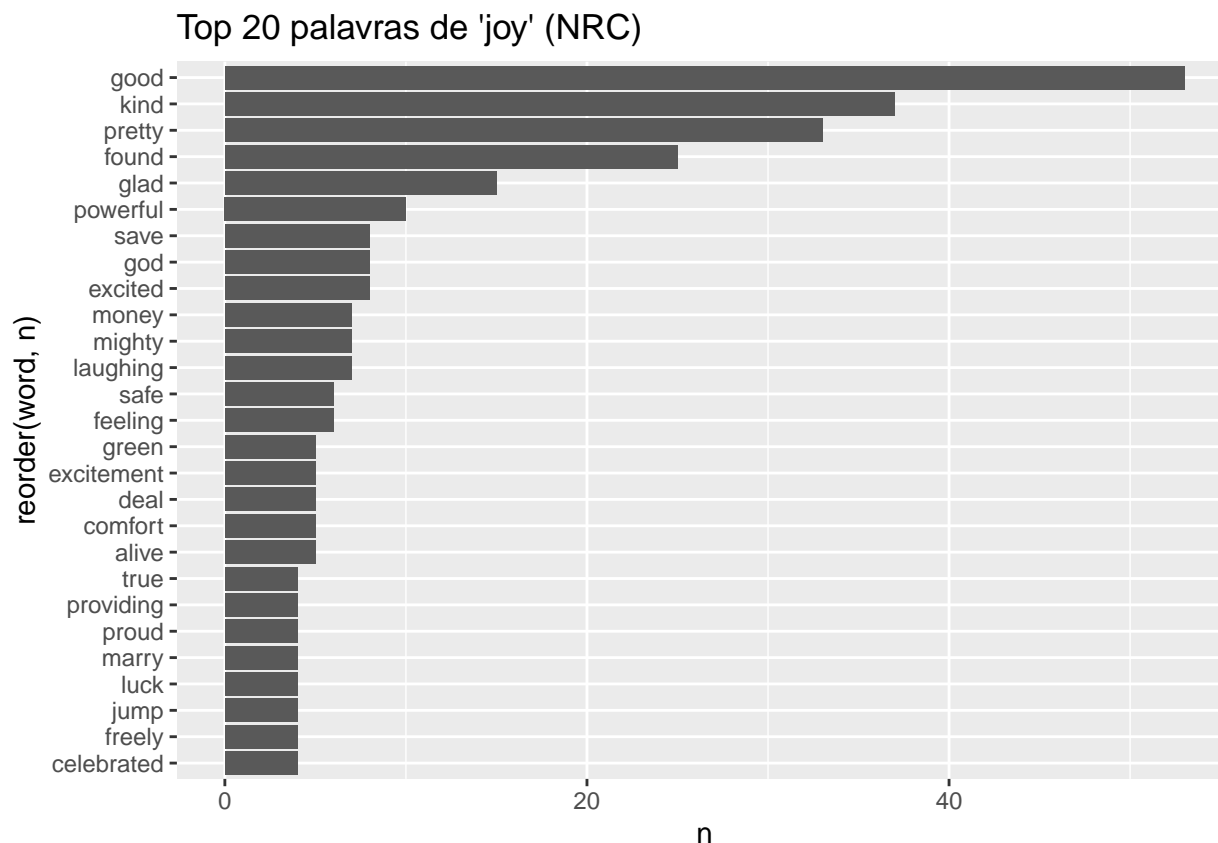
3. Saldo de Sentimento por Capítulo (Bing)

O gráfico de saldo por capítulo sintetiza a evolução emocional do enredo. Capítulos com saldo positivo destacam momentos de alegria, amizade e superação, como as cenas em que Tom e Huck encontram o tesouro ou participam de brincadeiras. Já capítulos com saldo negativo estão ligados a eventos de tensão, como punições, julgamentos ou situações de perigo. Essa visualização facilita a identificação dos pontos altos e baixos da narrativa, permitindo ao leitor correlacionar os resultados quantitativos com os principais eventos do livro. Por exemplo, é possível perceber que capítulos centrais do romance, onde ocorrem grandes reviravoltas, apresentam maior variação de sentimento, enquanto capítulos introdutórios ou de transição tendem a ser mais neutros. Isso reforça a coerência entre a análise de sentimentos e a estrutura literária da obra.



4. Top 20 Palavras de “Joy” (NRC)

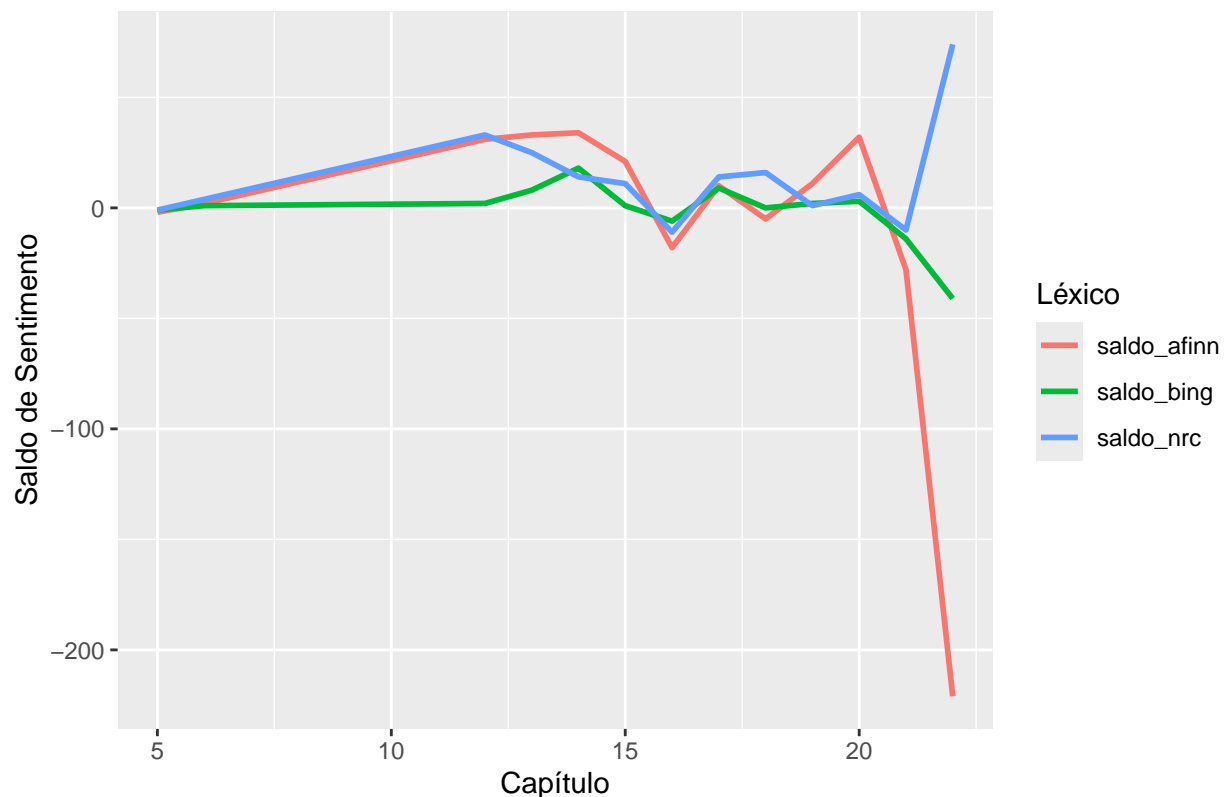
O gráfico das palavras de “joy” segundo o NRC revela como o léxico identifica os momentos felizes do texto. Termos como “play”, “friend”, “laugh”, “happy” e “freedom” aparecem com destaque, o que está alinhado com as cenas de brincadeiras, amizades e conquistas dos personagens. No entanto, é importante observar que o contexto pode modificar o significado dessas palavras: por exemplo, “play” pode aparecer em situações de conflito ou ironia, e “happy” pode ser usado em frases negativas. Isso mostra que, embora o NRC seja eficiente para capturar emoções explícitas, ele pode deixar de lado nuances contextuais, como felicidade implícita ou ironia. Uma análise mais profunda poderia cruzar essas palavras com trechos específicos do texto para validar se realmente representam momentos alegres.



5. Comparação de Saldos de Sentimento por Capítulo (Bing, NRC, AFINN)

O gráfico comparativo entre Bing, NRC e AFINN evidencia diferenças importantes na abordagem de cada léxico. O Bing apresenta oscilações mais claras entre positivo e negativo, sendo útil para uma visão geral do saldo emocional. O NRC permite analisar emoções específicas, como “joy”, “anger” ou “fear”, o que enriquece a interpretação dos capítulos. O AFINN, por sua vez, atribui pontuações de intensidade, capturando variações mais sutis de sentimento, mas pode ser menos sensível ao contexto literário. Ao comparar os três, nota-se que alguns capítulos são classificados de forma divergente, o que pode indicar limitações dos léxicos ou particularidades do texto. Essa análise reforça a importância de utilizar múltiplos métodos para obter uma visão mais completa e crítica da narrativa, além de estimular a reflexão sobre os limites das abordagens automáticas em literatura.

Comparação de Saldos de Sentimento por Capítulo



Quais capítulos apareceram mais positivos e mais negativos? Isso faz sentido na narrativa?

Os capítulos mais positivos, segundo os saldos dos léxicos, coincidem com momentos de aventura, amizade e superação, como os capítulos em que Tom e Huck encontram o tesouro ou participam de brincadeiras. Os capítulos mais negativos geralmente envolvem conflitos, punições ou situações de perigo, como o julgamento de Muff Potter ou o desaparecimento de Tom e Becky na caverna. Essa distribuição faz sentido, pois reflete os altos e baixos emocionais da narrativa, alinhando-se com os principais eventos do livro.

As palavras de joy capturadas pelo NRC representam bem os momentos felizes do livro?

O NRC identifica palavras associadas à alegria, como “play”, “friend”, “laugh”, “happy” e “freedom”, presentes nos trechos de brincadeiras e aventuras. No entanto, pode deixar de captar nuances contextuais, como ironia ou felicidade implícita, e pode incluir palavras em contextos neutros ou negativos, dependendo do uso no texto.

Entre Bing, NRC e AFINN, qual se mostrou mais adequado para esse texto? Justifique.

O Bing é mais direto, classificando palavras como positivas ou negativas, facilitando a visualização do saldo emocional por capítulo. O NRC é mais detalhado, permitindo analisar emoções específicas, mas pode gerar sobreposição de sentimentos. O AFINN atribui pontuações de intensidade, útil para medir variações sutis.

Para Tom Sawyer, o Bing é mais robusto para visão geral, o NRC é interessante para emoções específicas e o AFINN pode ser menos sensível ao contexto literário.

A ironia e o humor de Mark Twain foram bem capturados? Cite exemplos onde o algoritmo “erra”.

Os algoritmos de léxico têm dificuldade em captar ironia e humor. Frases em que Tom faz travessuras ou engana adultos podem ser classificadas como negativas devido ao uso de palavras como “trouble” ou “lie”, mesmo que o contexto seja cômico. O sarcasmo nos diálogos também pode ser interpretado erroneamente como sentimento negativo, mostrando que o modelo não entende o contexto narrativo, apenas o significado isolado das palavras.

Quais seriam as próximas melhorias (ex.: lematização, análise de negação, modelos para português/inglês contemporâneo)?

- Implementar lematização para agrupar variações de palavras (ex.: “play”, “playing”, “played”).
- Análise de negação para evitar erros em frases como “not happy” ou “never sad”.
- Utilizar modelos de sentimentos treinados para literatura ou inglês contemporâneo, que capturam melhor nuances de linguagem.
- Adotar abordagens de aprendizado de máquina ou redes neurais, que consideram contexto e sequência de palavras.
- Adaptar o pipeline para textos em português, caso o corpus seja traduzido ou comparado com obras nacionais.