

Introdução

Atualmente, a Betfair é a maior bolsa esportiva do mundo. Com mais de 15 anos em atividade e 4 milhões de usuários cadastrados, movimenta aproximadamente £400 milhões anualmente. Na Figura 1, podemos ver uma imagem do site.

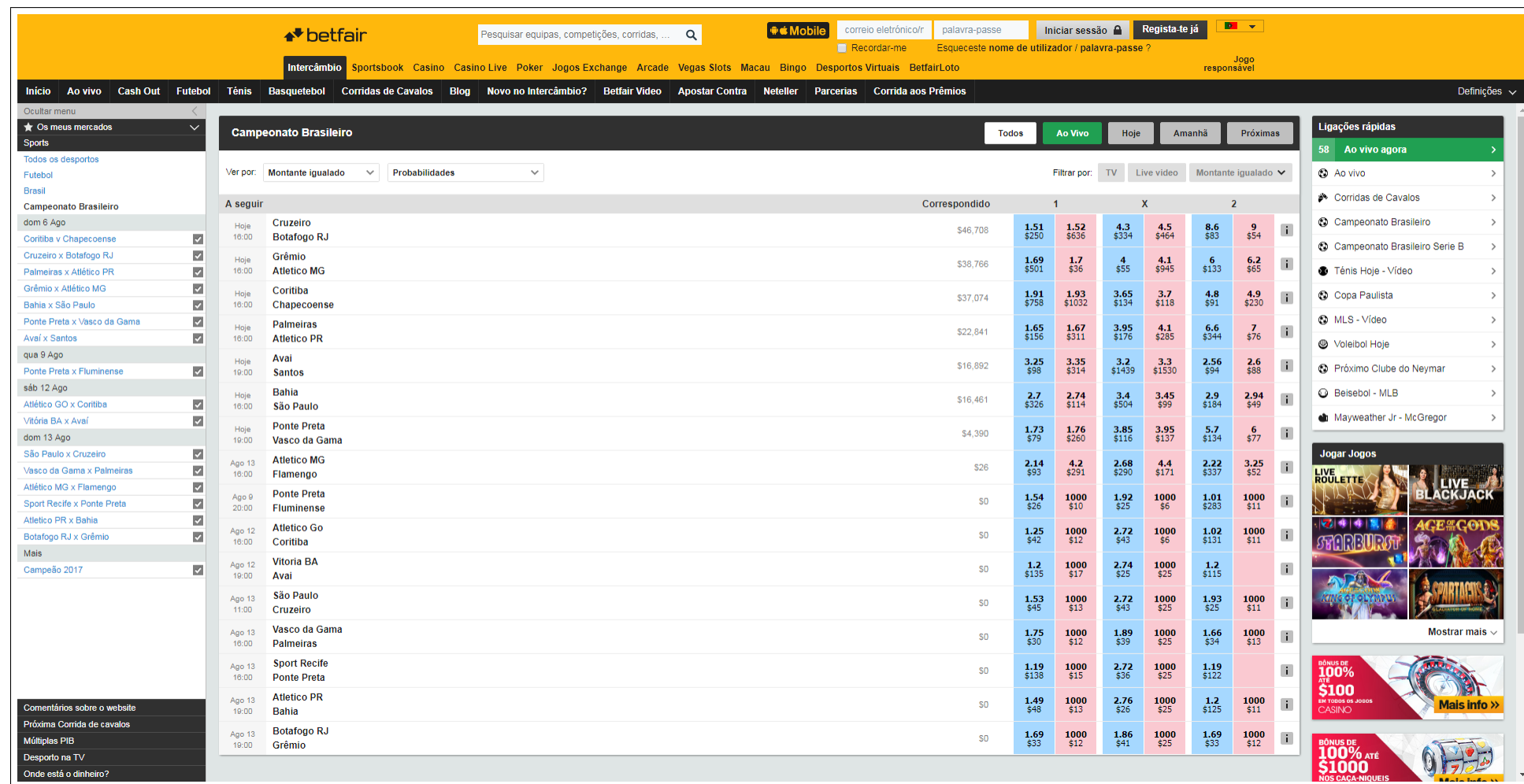


Figura 1: Imagem do site Betfair

Nesse projeto, buscamos a predição de qual categoria receberá um maior número de apostas em um período observado, seja o time da casa, time visitante ou opção de empate, analisando especificadamente os jogos da Série A do Campeonato Brasileiro de 2016.

Materiais e métodos

Para o desenvolvimento do projeto, foram tomadas uma sequência de etapas para o uso dos dados:

- Foram utilizadas a biblioteca *scikit-learn* para os algoritmos de aprendizado de máquina, e a biblioteca *NumPy* para manipulação de vetores e matrizes, ambas utilizando a linguagem Python.
- Também utilizamos a suíte *Weka* para aplicação dos algoritmos de aprendizado de máquina.
- Analisamos os dados oferecidos pela Betfair por todo o ano de 2016, em uma base de aproximadamente 55 milhões de linhas.
- Ao aglutinarmos os dados de uma mesma partida, atingimos o número de 379 linhas, a mesma quantidade de jogos no ano de 2016.
- Também foi necessário um pré-processamento dos dados oferecidos, como transformar datas e nome dos times em números (esse último via técnica de *one-hot encoding*), e adicionar a quantidade de vitórias de cada time nos últimos 5 jogos.
- No final de todas essas etapas e com todos os dados preparados para o processamento, finalmente aplicamos os algoritmos de aprendizado de máquina de categorização aos valores obtidos.

Resultados

Após filtrarmos e trabalharmos com os dados restantes, os primeiros resultados obtidos foram acerca da eficácia dos algoritmos preverem os resultados como um todo, exibidos na Figura 2:

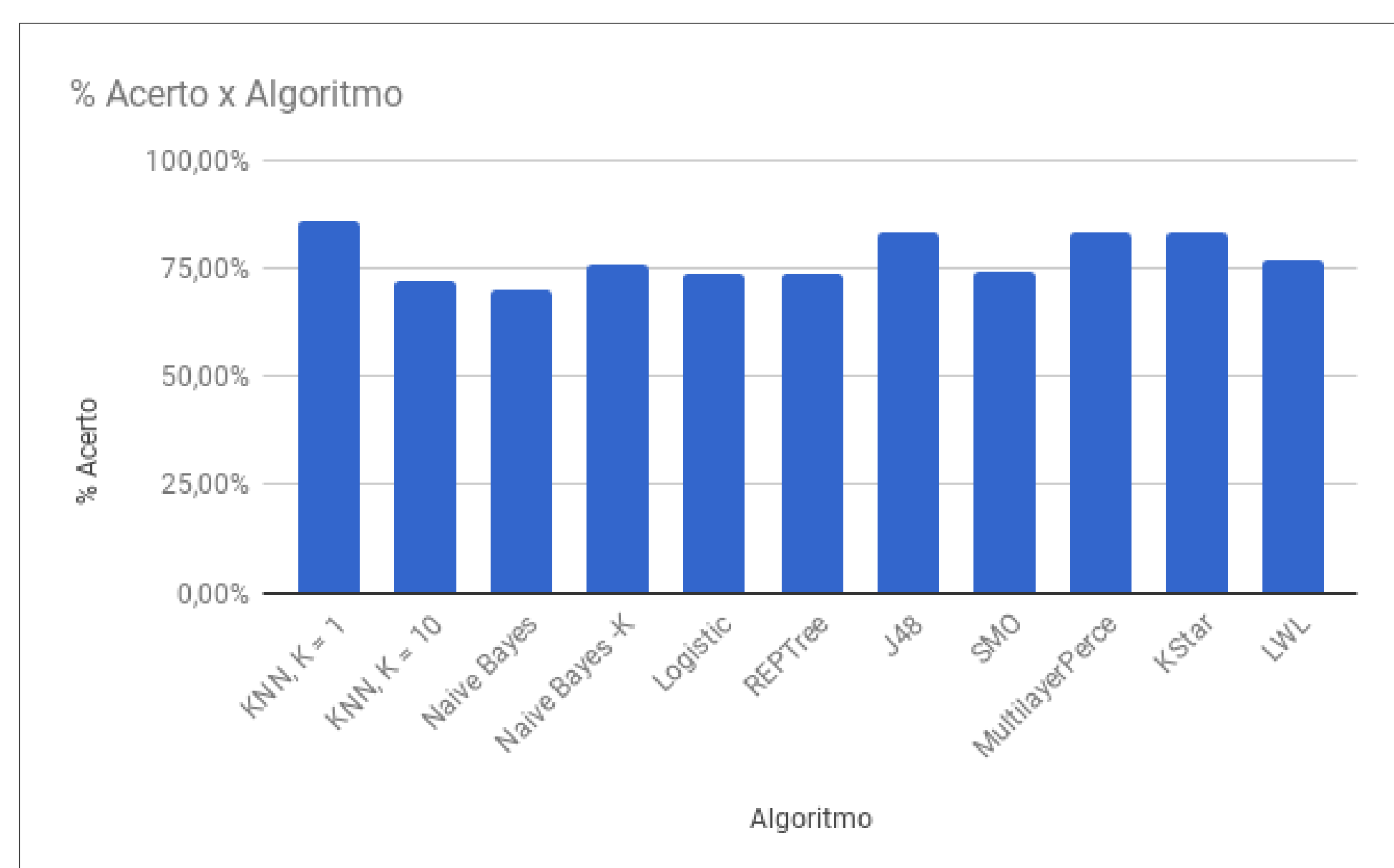


Figura 2: Porcentagem de acerto dos algoritmos para todos os resultados

Porém, esses resultados não são suficientes. Ainda que os algoritmos tenham bons resultados com todas as categorias, é necessário especificar a eficácia de cada algoritmo com cada categoria possível de apostas (time da casa, time visitante ou empate). Essa diferença se faz visível no próximo gráfico da Figura 3:

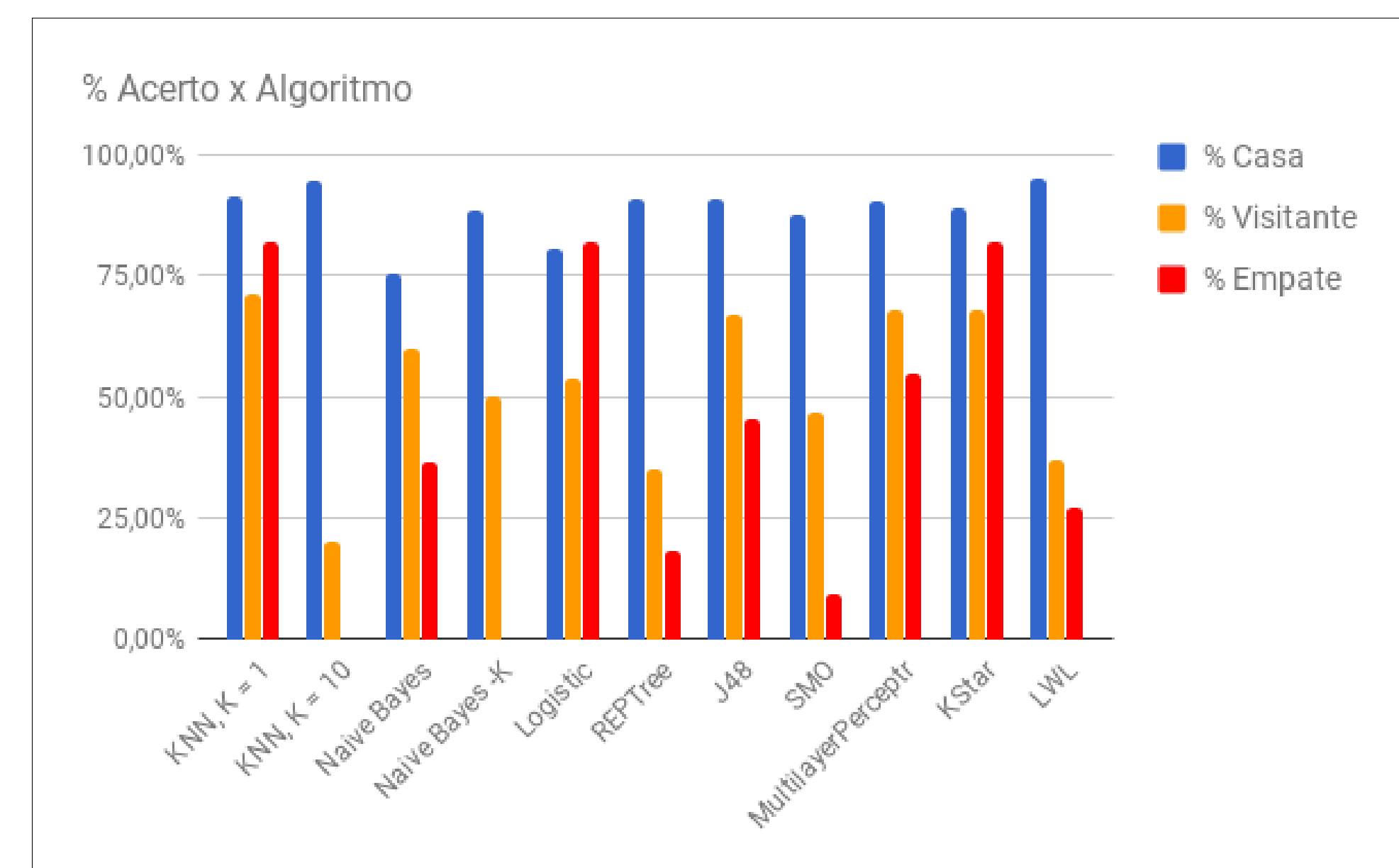


Figura 3: Porcentagem de acerto dos algoritmos para resultados de cada categoria

Outro valor que também podemos utilizar para análise é o chamado Coeficiente de Kappa, que mede a concordância entre os valores da classificação apresentada, apresentado na Figura 4:

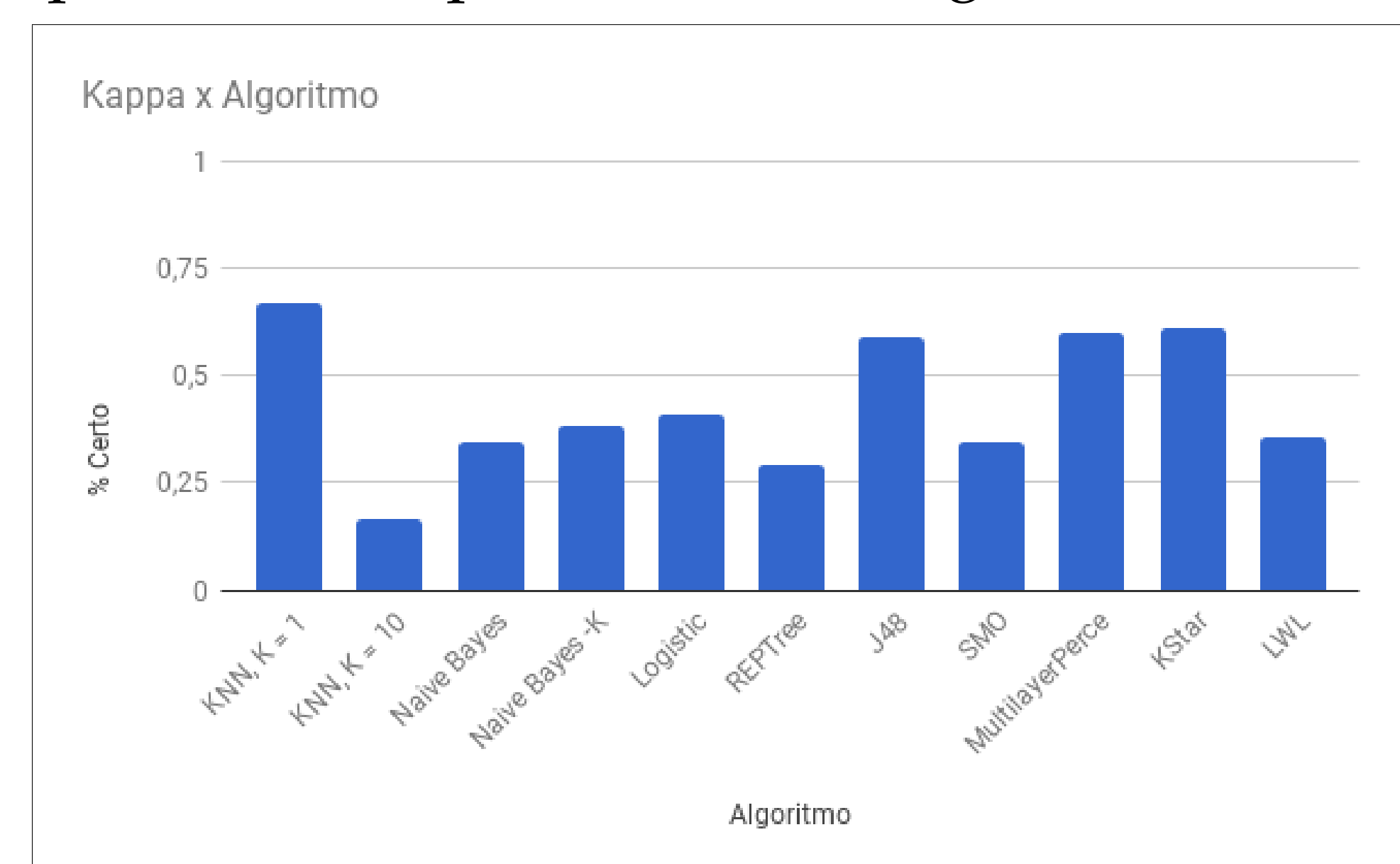


Figura 4: Porcentagem do Coeficiente de Kappa para cada algoritmo utilizado

Conclusões

Ao final dessas análises, fica visível que um valor alto de acerto para o total de dados não é necessariamente um resultado bem-sucedido, podendo ser apenas um falso positivo que ignora a situação das categorias em específico.

O próximo passo do nosso trabalho de predição será buscar a quantificação de apostas em cada opção das bolsas esportivas. Ou seja, iremos buscar quantas apostas cada opção receberá, ao invés de apenas indicar qual opção receberá mais apostas.

Como forma de buscar resultados mais eficazes, será necessário uma busca por bases mais completas em relação a eventos esportivos, com outros dados e estatísticas que possam melhor agregar no resultado final.

Agradecimentos

Por primeiro, gostaria de agradecer ao Instituto de Informática da Universidade Federal de Goiás, por oferecer o espaço para essa apresentação, além de ofertar essas e outras matérias relevantes para área.

Por segundo, gostaria de agradecer aos professores Anderson Soares e Gustavo Laureano, que ministraram a disciplina de Algoritmos de Reconhecimento de Padrões, e que junto com os outros colegas da turma, também ajudaram na construção dos trabalhos e no entendimento do conteúdo apresentado.

Por último, gostaria de agradecer a aqueles que reservaram um pouco do seu tempo para prestigiar essa apresentação.

Para maiores informações

- Email para contato: ayrtondenner_2013@hotmail.com
- Link para repositório no Github do poster: <https://github.com/ayrtondenner/ARP-2017-1/tree/master/Poster%20-%20Landscape>
- Seção da Betfair para jogos da Série A do Campeonato Brasileiro <https://www.betfair.com/exchange/plus/football/competition/13>
- Lista com dados históricos da Betfair: <http://data.betfair.com/>