

ABSTRACT

TOPIC DETECTION FOR CULINARY MARKET ANALYSIS USING LDA, NAÏVE BAYES ALGORITHM AND K NEAREST NEIGHBOR ALGORITHM

Muhammad Nurlazuardi Ajipawenang

14/360029/PA/15754

Food sales and market in general have become an important channel for consumption and with numerous and wide range of product selection, the variation and opportunity is quite vast. Despite its growth and general reach, understanding on this market are often limited to bias and speculation, the reason is that previously, there was a lack of documented data to be worked on. In recent years, social media has given quite a handful of opportunity to the consumer in terms of gauging the quality of the products by reading and examining the reviews posted by the users of online shopping platforms. Moreover, online platforms such as Amazon.com provides an option to the users to label a review as 'Helpful' if they find the content of the review valuable. This helps both consumers, merchants, and manufacturers to evaluate a general preferences in an efficient manner by monitoring the attribute that could indicate the sentiments on the reviews.

In this research, we offer a solution to assist the user, by using topic modeling that consists of LDA and topic coherence. Latent Dirichlet Allocation (LDA) is utilized to find the interrelated words related to heterogeneous products, also to find the most mentioned of words. Topic coherence is used to checks the value of topics to be comprehensible to human. By utilizing sentiment analysis, we could automatically assess the data regarding opinions, sentiments, and emotions expressed in communication. K-Nearest Neighbor (KNN) and Multinomial Naïve Bayes is used to classify the overall sentiment of the datasets. By using both topic modeling and sentiment analysis, this research will provide a way to incorporate information from social media to give further understanding of the market.

The result of this research shows that LDA topic modeling are able to identify and rank top words present in the dataset then the dataset could be visualized. The results of the LDA have been tested with coherence score to be semantically interpretable to human judgement. Coherence score are then calculated to the highest possible number of topics (1-49) 6 topics are chosen with the highest coherence score. While the sentiment analysis, in the form of KNN and Multinomial Naïve Bayes is capable to further analyze the dataset of its sentiment polarity with the results tends to be True Positive, hence providing additional insight of the dataset, through which further decision such as will this product will continue in the market or not.

Keywords: Food sales and market, LDA, Sentiment analysis, KNN, Multinomial Naïve Bayes.

DETEKSI TOPIK UNTUK ANALISIS PASAR MAKANAN MENGGUNAKAN ALGORITMA LDA, NAÏVE BAYES, DAN K-NEAREST NEIGHBOR

Muhammad Nurlazuardi Ajipawenang

14/360029/PA/15754

Pada umumnya, penjualan dan pasar makanan adalah lingkup konsumsi yang penting dan dengan berbagai macam pilihan produk, variasi dan kesempatan yang ada di dalamnya cukup luas. Walaupun memiliki pertumbuhan dan jangkauan yang luas, ilmu mengenai pengertian pasar ini seringkali terbatas dengan bias dan spekulasi, karena sebelumnya, ada kekurangan dokumentasi data yang bisa diproses dan pelajari. Di tahun-tahun belakangan ini, media social telah memberikan kesempatan untuk konsumen mengetahui dan mengukur kualitas dari produk yang ada dengan cara membaca dan memeriksa ulasan yang diunggah oleh pengguna platform belanja online. Ditambah lagi, platform online seperti Amazon.com menyajikan pilihan kepada pengguna untuk menandai sebuah ulasan sebagai “berguna” apabila ulasan tersebut dianggap berisi dan berharga. Hal ini sangat membantu konsumen, pedagang, dan pabrik untuk secara efisien mengevaluasi preferensi umum dengan cara memindai atribut yang mengindikasikan sentiment pada ulasan.

Dalam penelitian ini, penulis menyuguhkan solusi untuk membantu pengguna, dengan menggunakan topic modeling yang memanfaatkan algoritma LDA dan nilai topic coherence. Latent Dirichlet Allocation (LDA) digunakan untuk mencari kata-kata yang berhubungan dengan produk yang heterogen, juga untuk mencari kata yang sering digunakan. Topic coherence digunakan untuk memeriksa nilai dari topik yang bisa dimengerti oleh manusia. Dengan menggunakan analisis sentiment, penulis bisa menggunakan data opini, sentiment, dan emosi yang diekspresikan didalam komunikasi online. K-Nearest Neighbor (KNN) dan Multinomial Naïve Bayes digunakan untuk mengklasifikasi sentiment seluruh dataset. dengan menggunakan topic modeling dan analisis sentiment, penelitian ini berupaya memberikan cara untuk menggabungkan informasi dari media sosial untuk memahami lebih dalam segmentasi pasar.

Hasil dari penelitian ini menunjukkan bahwa topik model LDA mampu mengidentifikasi dan menggolongkan kata-kata utama dari dataset, lalu divisualisasikan. Hasil dari LDA telah diuji dengan nilai topic coherence untuk memeriksa apakah topik tersebut dapat dipahami manusia. Dengan nilai topic coherence tertinggi (dari 1-49) 6 topik dipilih dengan nilai coherence tertinggi. Dengan analisis sentiment, menggunakan KNN dan Multinomial Naïve Bayes dataset dipahami lebih jauh seputar perbedaan sentiment yang ada, dengan hasil yang menunjukkan kecenderungan nilai menjadi True Positive, yang mana memberikan pengertian lebih tentang dataset tersebut, dan mampu membantu pemilihan keputusan tentang produk di pasar tersebut.

Kata Kunci: Penjualan dan pasar makanan, LDA, analisis sentiment, KNN, Multinomial Naïve Bayes.