

Opinion Mining for Mcdonald's Services Using Machine Learning

Sujatha R*, M Sai Praneeth and M Lakshmi Deepak

School of Information Technology and Engineering, VIT University, Vellore

*Corresponding author: E-Mail: r.sujatha@vit.ac.in

ABSTRACT

The work is carried on opinion mining for McDonald's services using machine learning. The people's interest to provide the feedback and review has been uptrend in this modern world. Based on the opinions and reviews inputted by the people, the sentiment classifier classifies the emotion and polarity levels of the review. The opinion mining results helps to make the better decision. For this purpose the opinion of McDonald's reviews extracted from Twitter API and TripAdvisor API. For the classification of sentiment purpose the Naïve Bayes, SVM and Decision Tree algorithm are utilised. With the help of these algorithms the polarity levels of the opinion are classified as positive, negative and neutral. In addition the Reverse Geocoding algorithm and Google map API mechanism are employed such that it benefits the user to find the best and nearest McDonald's branch that will display based on the results of opinion mining polarity results.

KEY WORDS: Opinion Mining, Machine Learning, Naive Bayes, SVM, Reverse Geocoding, Google Map API, Sentiment Classification, polarity, sentiment.

1. INTRODUCTION

In the modern era the reviews of people plays a vital role. It is complex to analyse people's reviews, attitudes, opinions, and sentiments towards a product, organization, service, institute, event, item or other issues. To address this issue the concept named sentiment Analysis or opinion mining came into existence. Sentiment analysis is analysing the opinions, sentiments and emotions of people's or user's which are in the format of text or sentence or document to define the polarity levels. Sentiment analysis applied in almost every business because opinion's plays a key role for all human activities.

Purpose of this is to sympathize how the opinion mining for McDonald's services (OPMS) by using various machine learning techniques system has developed. System provides the nearest and polarity level of McDonald's branches according to city selected by user. The polarity level defined by opinion mining which helps the user's to know the positive, negative and neutral percentages of particular branch.

The major contributions of the work in developing the OPMS are as follows:

- Extraction of McDonald's TripAdvisor review dataset using TripAdvisor API and Twitter using twitters API.
- A detailed study of ML techniques for opinion mining of McD reviews.
- Visualization of the sentiment results to the users with the spatial data.

Literature Survey: An application for smart hospitalization had been created, that uses the sentiment classification results of the doctor and hospital reviews. This was done by using sentiment algorithm. They used reverse geocoding algorithm to find the nearest hospital corresponding the input spatial location of the patient. The reverse geocoding concept is added to the proposed system (OPMS) which will provide the users with the nearest and best McDonald's branch.

The work is carried on based on the aspects of opinion mining. They proposed a different approach for opinion mining which uses the word dictionary Senti Word Net to find the sentence level feature score.

The researchers tested the effectiveness of the machine learning techniques to the sentiment classification issue. The sentiment classification is mainly done by taking the keywords like good, worst, not good, happy etc. They raised an issue of having no single keyword in the sentence that will become negative. So, they analysed these difficulties to get better accurate results. For increasing accuracy in classification they used three different machine learning algorithms.

The survey on opinion mining and sentiment analysis was carried out. For different types of datasets they applied machine learning approach with different supervised learning algorithms. They proposed Naive Bayes, SVM, KNN and Maximum Entropy for the sentiment classification of the reviews. They compared the performance of these algorithms.

The similar area of research which they applied opinion mining on different travel review sites like TripAdvisor and Booking.com. They had calculated the accuracy of opinion mining using machine learning and sentiwordnet. They had implemented the visualization of sentiment results on Google Maps. Based on the hotel reviews whether it is 'worthy' or 'bad' will be visualized on the Google Maps. The survey on various machine learning techniques and explained the concept of Kernel SVM.

The detailed survey on the various data sources and the previous works on sentiment analysis is carried on. They raised some challenges on sentiment classification of reviews with orthographic words, typographical mistakes, abbreviations etc.

The major work used sentiment analysis algorithms such as SVM and KNN. They proposed the combination of both classification and clustering concepts for the increase of accuracy. The researchers had compared the three machine learning algorithms for the movie reviews dataset. In their research the SVM model outperformed than the other algorithms. The work is performed related to opinion mining on the student's feedback on examination, teacher dataset.

The researchers proposed a method of integrating combination of different types of data, assessing them and visualizing on the interactive maps. This method is very useful for the business planners and scientists for strategic planning of decisions. The work discusses about the various techniques to sentiment analysis, mainly machine learning and some cognitive approaches. They mainly explained about how to convert the unstructured data into structured data.

The researchers proposed a system that investigates the issue of analysing public sentiment variations and finding the possible reasons which causes the variations. The work discussed the challenges like classification of the non-word tokens, emoticons in reviews. In their research the Naive Bayes model outperformed than the SVM.

The main aim is to design a new tool which determines the parts of speech opinion mining words using the sentiment classification technique.

Existing System: Every activity of humans mostly depends on other's opinions. We primarily consider people's opinion to take a decision over any things like product, item or service.

For example to buy a mobile we always consider other experience and then we take decision over that product. But, in present days everyone are interested to use web applications to sell or purchase the goods. This web applications provide comment or review column to share the experience of users about particular product which they have brought. By analysing this reviews user can proceed to make a decision over that particular product. But manually analysis over a service or product becomes difficult in the case of large set of reviews. So it is hard for users to make a conclusion over a service or product and companies can't get accurate intension of users over a product. This situation can be handled or setback by opinion mining.

2. PROPOSED SYSTEM

The architecture of the system is given in Figure.1. The following modules describe each sequent steps in detail. All the phases of the system was implemented using R as a programming language and RSTUDIO as a development IDE.

Data Collection: This module deals with the collection of opinions or reviews of various McDonald's branches from Twitter and TripAdvisor API. They have huge volume of McD reviews data. The reviews of customers are extracted for each branch and stored in separate document. This is the major step in the process of sentiment classification. These documents are used for next steps such as Pre-processing, Classification and for Visualization of results.

Pre-Processing: Data extracted from blogs and tweets contains large amount of unnecessary data like abbreviations, shortcuts, numerical values, #tags, emoji's and various symbols. But this type of data will be removed by Pre-processing.

Pre-processing of opinion or reviews supports to develop the accuracy of opinion mining and it also helps to fast classification of polarity's and emotions.

Pre-processing leads to:

- Convert input document to plain text
#Happy New Year\$@4Love You happy new year love you
- Removal of unnecessary characters, stop words, stemming and punctuations.
- The major challenge is the conversion of orthographic words into general terms. For example: a) "gud", "gd" rather than "good".

"McD" rather than "McDonald's".

Sentiment Classification: Sentiment Classification is mainly consists of two tasks which includes:

- Sentiment Polarity Classification.
- Sentiment Emotion Classification.

The sentiment polarity assignment is a sentiment classification task which deals with analysing the text and gives the polarity results like positive, negative or neutral orientation. Sentiment emotion deals with analysing emotion of the text and gives emotion classification results like Joy, anger, disgust, fear, sadness and unknown. There are many procedures and tasks to achieve sentiment analysis. To achieve the goal of sentiment analysis we have to find the polarity classification and emotion detection of the particular data target. In order to achieve sentiment analysis we have used different types of tools and techniques.

Machine Learning approach is used to find the sentiment classification of the McDonald's services. The machine learning approach is the best technique for the sentiment classification of the reviews.

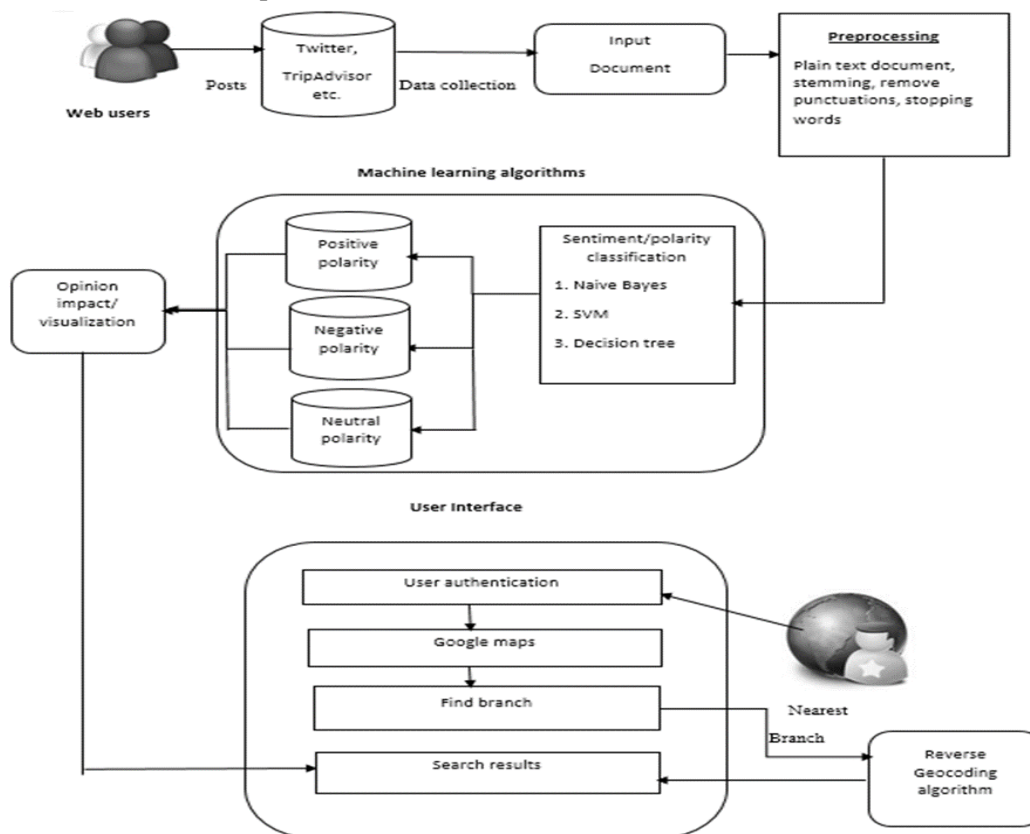


Figure.1. Architecture diagram
Table.1. Pre Processing

Pre-processing Type	Before	After	Description
Stop words	a) I hate this mc donald's service b) I am very happy with this project	a) Remained Word's hate, mc donald's, service Removed Word's I and This are stop words b) Remained Word's very, happy, project Removed Word's I, am, with and this	Stop words are used in sentences just to convey message in correct way. But they do not useful in classifying emotion or polarity.
Stemming	a) Loves, Loving, Loved b) Winner, Winning, Won c) The student's mobiles are different colours	a) Love b) Win c) The student mobile are differ colour.	Stemming usually refer to chops off the ends of words in order to achieving accuracy. In general various verbs forms are used for grammatical reason in sentences. But verbs forms does not play any key role in text analysis of review.
Punctuations, Case and hyperlinks Conversion	It's very good Link	It's very good link	It removes all the punctuations and hyperlinks in giving input. To make result more efficient.

Machine learning methods: The main aim of Machine Learning is to build a classifier which classifies the sentiment of a trained data. To classify the sentiment the major tasks are divided into sub-tasks: Pre-processing of data; Sentiment Classification; Post Processing; Visualization.

Sentiment Classification is the major level of the process which finds the mapping between patterns and targets. To classify the sentiment of McDonald's TripAdvisor and Twitter reviews we used some Machine Learning approaches. They are: Naive Bayes Classification; Support Vector Machines; Decision Trees.

Naive Bayes Classification: This approach is best suitable for the text classification. For the review document 'b' the classifier will assign the class label as $a^* = \operatorname{argmax}_c P(a|b)$. This probabilistic classifier is dependent on the Bayes theorem. The NB Classifier is more suited when the proportions of inputs are high. The Bayes rule used by the NB Classifier,

$$P(a|b) = \frac{P(a) P(b|a)}{P(b)}$$

Where $P(a)$ and $P(b)$ be the probabilities of the perceiving a and b and $P(a|b)$ is the probability of perceiving event a given that event b is true and $P(b|a)$ is vice-versa of $P(a|b)$.

Support Vector Machines: Support vector machines (SVM) can be assumed as a top of the level that defines a boundary between different points of text data. The main aim of the SVM is to create a Hyperplane which leads to the partition of data on both the sides.

Here we discuss about the SVM kernels for non-linear spaces. In many applications, the relationship between the various variables will be non-linear in nature. To avoid this problem of non-linearity SVM provides kernel process which has an ability to map the non-linearity problem into a higher dimension space.

The system has a nonlinear relationship between the two features: emotion and polarity.

The kernel function is of the form

$$K(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j)$$

Where $\phi(x)$ is the mapping of text data into another space.

Decision Tree: Decision tree classifier provides an ordered decomposition of the training dataset in which a business rule on the attribute data is used to break the records or data. The division of dataset is done iteratively until the leaf nodes incorporate minimum number of tuples or records which are used for the purpose of classification.

For S samples of training data set the expected information or entropy needed to classify a tuple or record in the dataset D is:

$$I(S_1, S_2, \dots, S_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

Where p_i is the probability that an arbitrary sample belongs to class label v_i . The entropy correlated with using attribute A as the test attribute is assigned as:

$$E(A) = \sum_{j=1}^m \frac{S_{1j} + \dots + S_{mj}}{S} \cdot I(S_{1j}, \dots, S_{mj})$$

After applying the decision tree algorithm for the McDonald's reviews the result obtained is in Fig.2.

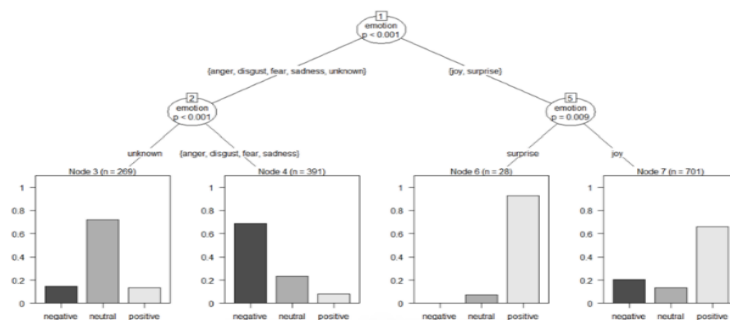


Figure.2. Decision tree result

Reverse geocoding: The Reverse Geocoding algorithm is the process of converting geographic coordinates into a human understandable address. The Google maps Geocoding API will find the location for the given place Id and it also provides direct access via HTTP request. The Euclidean distance formula is utilized for finding nearest McD branch.

3. RESULTS AND ANALYSIS

Since, we used three algorithms Naive Bayes, SVM and Decision Tree for the classification of sentiment analysis the SVM model has outperformed than the other classifiers.

After applying the NB classifier to the McDonald's reviews the obtained polarity results positive, negative and neutral which are in Figure.3.

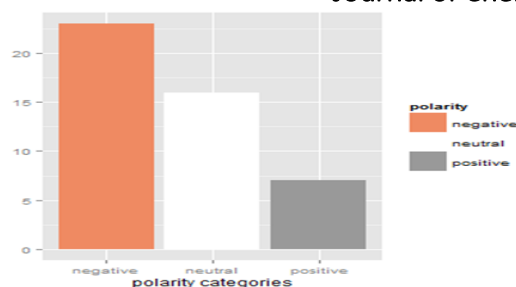


Figure.3. Polarity results

From the above graph we analyse that the McDonald's reviews taken from the TripAdvisor site is having high negative polarity and low positive polarity reviews.

The Figure.4, is the Emotion Classification results for the McDonald's reviews using NB classifier That contains high negative emotions (anger, fear, disgust, sadness) and low positive emotion (Joy, surprise) reviews.

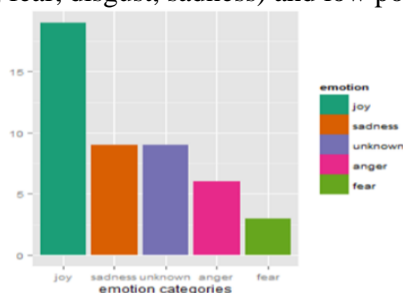


Figure.4. Emotion results

4. CONCLUSION

By using opinion mining concept for any organisation or industrial services there would be improved customer satisfaction, efficiency, inventory, improves the quality of service. The user gets the recommended information based on the previous user's opinions of the service to make the correct decisions. The proposed system has a facility for the user to find a best and nearest McDonald's branch. To obtain this the opinion mining concept is used. Using that we have classified the reviews and found the polarity and emotion results. It would be very beneficent to the users to find the best branch. This would help in suggesting the users to make a best decision on different branches.

Future Scope: We proposed a service to the user for finding nearest branch which is having high positive polarity results. The application is very user friendly by collecting reviews on the McD branches from various sources and providing the best results of the branch and its location to the user by the help of opinion mining and Google Maps API.

In this system only word sense disambiguation alone considered and some explicit features are ignored. Next works aims to incorporate the analysis of orthographic words in the pre-processing phase and also analysis of sentences like conditional which will increase the accuracy of the algorithms.

REFERENCES

- Bjorkelund, Eivind, Thomas H, Burnett, and Kjetil Norvag, A study of opinion mining and visualization of hotel reviews, Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, ACM, 2012.
- Chinsha T.C and Shibily Joseph, Aspect based Opinion Mining from Restaurant Reviews, IJCA Proceedings on International Conference on Advanced Computing and Communication Techniques for High Performance Applications, Foundation of Computer Science (FCS), 2015.
- Dhanalakshmi V, Dhivya Bino and Saravanan A.M, Opinion mining from student feedback data using supervised learning algorithms, 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), IEEE, 2016.
- Dhokrat, Asmita, Review on Techniques and Tools used for Opinion Mining, International Journal of Computer Applications Technology and Research, 2015, 419-424.
- Gayatri Datre, Yograj Deore, Akash Jadhav, Kunal, Smart Hospitalization and Patient Review Analysis, International Journal for Scientific Research and Development, 2016, 1267 - 1270.
- Kalaivani P and Shunmuganathan K.L, Sentiment classification of movie reviews by supervised machine learning approaches, Indian Journal of Computer Science and Engineering, 2013, 285-292.

Kasper, Walter, and Mihaela Vela, Sentiment analysis for hotel reviews, Computational linguistics-applications conference, 231527, 2011.

Khairnar, Jayashri, and Mayura Kinikar, Machine learning algorithms for opinion mining and sentiment classification, International Journal of Scientific and Research Publications, 3.6, 2013, 1-6.

Kodipaka, Rajeshwar Rao, Sanjeeva Polepaka, and Md Rafeeq, Design of Sentiment Analysis System using Polarity Classification Technique, International Journal of Computer Applications, 2015, 22-24.

More, Kalyani, Survey of Evaluating Public Opinion Variation on Twitter, International Engineering Research Journal, 2016, 3532-3535.

Mulay, Snehal A, Sentiment Analysis and Opinion Mining With Social Networking for Predicting Box Office Collection of Movie, International Journal of Emerging Research in Management & Technology, 2016, 74-79.

Pang Bo, Lillian Lee and Shivakumar Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Volume 10, Association for Computational Linguistics, 2002.

Pino, Carmelo, Isaak Kavasidis and Concetto Spampinato, Assessment and visualization of geographically distributed event related sentiments by mining social networks and news, 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC) IEEE, 2016.

Sneka G, Vidhya CT, Algorithms for Opinion Mining and Sentiment Analysis, An Overview, International Journal of Advanced Research in Computer Science and Software Engineering, 2016, 455-459.

Vinodhini G, and R.M, Chandrasekaran, Sentiment analysis and opinion mining, a survey, International Journal of Advanced Research in Computer Science and Software Engineering, 2012, 282-292.