International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 312-313
Published Online May 2020 in IJEAST (http://www.ijeast.com)

FINANCIAL TRACKER USING NLP

Deepak Jain, Dishi Singh, Astha Gupta, Agraj Singh
Department of CSE
IMS Engineering College
Ghaziabad, Uttar Pradesh, India

Mrs. Shruti Keshari
Department of CSE
IMS Engineering College
Ghaziabad, Uttar Pradesh, India

***Abstract*— NLP (Natural Language Processing) is a tool that helps computers to understand natural languages like english. In general, computers can understand data, tables etc. which are well structured. But when it comes to natural languages, it's not possible for computers to interpret them. NLP helps to convert the natural language in such a manner that can be easily processed by modern computers. Financial Tracker is an approach that will use NLP as a tool and will classify the user messages in various categories. The application of the approach can be seen at many levels. At an individual level, this allows us users to filter out useful financial messages from a large junk of messages. On the other hand, from an organisation point of view, this is useful in services like online loan disbursal, which are hitting the market nowadays. These services try to provide online loans to individuals in a quick manner. But when It comes to business point of view, loan recovery from customers becomes a very much crucial aspect. As most such services actually can't take strict legal actions against the fraud customers, it becomes a requirement that loan should be provided only to those who deserve it. At that point this model can come under picture. As a business we can find the user's messages from their inbox (after taking permission from the users). These messages can be filtered using NLP which can help to differentiate various kinds of messages in the user's inbox which can further be used as a knowledge base for further prediction on user's behaviour in terms of money related transactions.**

Keywords— **Supervised learning, Text classification, Machine Learning, Classification, Regex**

I. INTRODUCTION

Natural Language Processing has become an important application of machine learning. Machine learning is traditionally very famous for its prediction and classification algorithms. NLP also uses these features of machine learning to gain insights on textual data, build correlations etc which is used to process the human language efficiently. According to various estimations, very less of the available data is present in structured form. We all are surrounded by data. Data is being generated as we speak, as we tweet, as we send various messages on whatsapp and in various other activities. Majority of this data exists in the textual form, which is highly unstructured in nature. Even if we have high dimensional data, the information present in it can't be directly used unless it is processed (read and understood) manually or analyzed by a computer system. The approach discussed here is used to solve this problem of text data processing by automated computers and use the results for prediction. The cibil score generator is an idea to use NLP in user's messages processing which can be used by organisations that provide online loans to the customers who don't have any credit history and hence no cibil score data. This method can help to generate an estimated cibil score with the help of a preprocessed NLP model, which classifies user's inbox messages as credit-amount, debit-amount and recharge-amount messages which can help to predict the monthly expenditure, earning and lifestyle of the individual.

II. EXPERIMENT AND RESULT

In this research, we have collected sample messages for building raw data that was preprocessed for further actions. The filtration process involves following steps

- Text Transformation to lower case letters.
- Removal of unwanted symbols, stopwords from text using regex. These stopwords include words like ("the", "a", "an", "in"), which can be ignored easily with very little or no loss of information.
- Stemming - It stems the words to find the root word. For example a set of words (classifier,classifies,classify) will be stemmed to the word classify.

All the above steps are part of NLP data pre-processing. All these steps make the data lighter and less noisy. Then, the data was converted to the matrix which contains all the different words in the data as the attribute name and each sentence (or text message) as a row. Each cell contains either a zero or one based on the presence of the word in the specific sentence. We have used countvectorizer from sklearn for this task. The data is now in the form which can be processed by a computer system. Then, resampling is done to resample various kinds of messages to handle multiple output class imbalance in the data. Then with the help of random forest classification algorithms from machine learning, various messages have been classified in specified categories for further analysis. Random forest, as the name implies, consists of a large number of individual decision trees that work as an ensemble (collection of decision trees). Each decision tree in the random forest brings out a class prediction and the class with the most votes becomes the model's prediction.

After Message classification, different messages are searched for the message genre which helps to predict monthly expenditure and earnings of an individual at a single place.

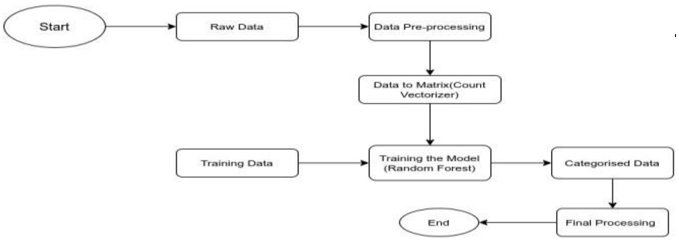

Fig. 1. Flowchart

III. CONCLUSION

In this study, we have found a very useful application of NLP from text messages. Such a knowledge base and data can be used for an enormous number of studies in the field of NLP. This model can be very useful in sectors like banking, e-commerce along with its significance in the personal life for budget management.

ACKNOWLEDGEMENT

My sincere thanks to guide Mrs. Shruti Keshari, Professor IMS Engineering College, Ghaziabad, India for her patient guidance and productive suggestions for the research in the area of financial tracker using natural language processing. This research is made possible due to her willingness to devout time graciously.

IV. REFERENCES

[1] Ingrid E. Fisher, Margaret R. Garnsey, Mark E. Hughes, 2016, "Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research",10.1002/isaf.1386.

[2] Venera Arnaoudova, Sonia Haiduc, Andrian Marcus, Giuliano Antoniol , 2015, "The Use of Text Retrieval and Natural Language Processing in Software Engineering", 10.1145/2889160.2891053.

[3] Tushar Ghorpade, Lata Ragha, 2012, "Featured based sentiment classification for hotel reviews using NLP and Bayesian classification", 10.1109/ICCICT.2012.6398136.

[4] Monisha Kanakaraj, Ram Mohana Reddy Guddeti, 2015, "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques", 10.1109/ICOSC.2015.7050801

[5] Shweta C. Dharmadhikari, Maya Ingle, Parag Kulkarni, 2011, "Empirical Studies on Machine Learning Based Text Classification Algorithms", 10.5121/acij.2011.2615.

[6] Gobinda G. Chowdhury, 2005, "Natural Language Processing", 10.1002/aris.1440370103

[7] Erik Cambria, Bebo White, 2014, "A Review of Natural Language Processing Research", 10.1109/MCI.2014.2307227

[8] Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman, 2011, "Natural language processing: an introduction", 10.1136/amiajnl-2011-000464.

[9] Rui Xia, Chengqing Zong, Shoushan Li, 2011, "Ensemble of feature sets and classification algorithms for sentiment classification, 10.1016/j.ins.2010.11.023.

[10] Atefeh Farzindar, Diana Inkpen, 2015, "Natural Language Processing for Social Media", 10.2200/S00659ED1V01Y201508HLT030

[11] E. Stamatatos, N. Fakotakis, G. Kokkinakis, 2000, "Text genre detection using common word frequencies", 10.3115/992730.992763

[12] Ellen Riloff, Wendy Lehnert, 1994, "Information extraction as a basis for high-precision text classification", 10.1145/183422.183428

[13] Teresa Gonçalves, Paulo Quaresma, 2004, "The impact of NLP techniques in the multilabel text classification problem", 10.1007/978-3-540-39985-8_46.

[14] Andronicus A. Akinyelu , Aderemi O. Adewumi, 2014, "Classification of Phishing Email Using Random Forest Machine Learning Technique", 10.1155/2014/425731

[15] Olga Ormandjieva, Ishrar Hussain, Leila Kosseim, 2007, "Toward a text classification system for the quality assessment of software requirements written in natural language", 10.1145/1295074.1295082.