THE ENGLISH LANGUAGE MULTILINGUAL PROCESSING FOR SENTIMENT ANALYSIS IN SOCIAL MEDIA BY USING PYTHON NLTK TEXT CLASSIFICATION, MIOPIA AND MEANINGCLOUD SENTIMENT ANALYSIS TECHNIQUES

ALLEN LEE WEI KIAT

BACHELOR OF COMPUTER SCIENCE

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science (Computer Systems & Networking) With Honours.

(Supervisor's Signature)

Full Name : DR. NOR SARADATUL AKMAR BINTI ZULKIFLI

Position : SENIOR LECTURER

Date : 13 MAY 2019

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Student's Signature)

Full Name : ALLEN LEE WEI KIAT

ID Number : CA15088

Date : 13 MAY 2019

THE ENGLISH LANGUAGE MULTILINGUAL PROCESSING FOR SENTIMENT ANALYSIS IN SOCIAL MEDIA BY USING PYTHON NLTK TEXT CLASSIFICATION, MIOPIA AND MEANINGCLOUD SENTIMENT ANALYIS TECHNIQUES.

ALLEN LEE WEI KIAT

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science (Computer Systems & Networking) With Honours

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

MAY 2019

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest appreciation to all those who provides me the possibility to complete the PSM report. A special gratitude I would like to give to my supervisor Dr. Md Manjur Ahmed. Without his guidance on doing the PSM report, I cannot finish this PSM report successfully. He gives encouragement and contribution to me throughout the research. He gave me inspiration and advices constantly when doing the thesis. Besides that, I would like to thank my supervisor, Dr. Nor Saradatul Akmar Binti Zulkifli, for the advice and encouragement she has provided. I have been quite lucky to have a supervisor who cared so much about my work(thesis). She responded to my queries and questions promptly.

Furthermore, I also want to give my sincere appreciation to the UMP Faculty of Computer Systems & Software Engineering which always supporting all the students who are taking the PSM. Besides that, I would like to thank my family and friends that always give fully support and good advices.

ABSTRAK

Pada masa kini, kebanyakan syarikat telah menggunakan Internet untuk menawarkan perkhidmatan dan produk mereka. Pelanggan Internet dapat melihat ulasan pelanggan-pelanggan terhadap produk atau perkhidmatan sebelum mereka hendak memilih untuk membeli sesuatu barang atau menonton filem. Syarikat perlu menganalisis sentimen dan perasaan pelanggan berdasarkan ulasan mereka. Hasil analisa sentimen menjadikan syarikat mudah untuk mencari ungkapan pengguna mereka adalah lebih positif atau negatif. Analisis sentimen digunakan dalam *data mining*. Ketepatan hasilnya adalah isu analisis sentimen. Objektif penyelidikan ini adalah untuk meneroka dan menilai bahasa Inggeris dengan menggunakan 3 teknik analitik sentimen yang berbeza iaitu *Python NLTK Text Classification, Miopia dan MeaningCloud* dari segi klasifikasi teks mereka (positif atau negatif). 3 teknik analisis sentimen yang telah digunakan dalam kajian ini untuk menganalisis analisis sentimen ulasan dan ulasan dari bahasa Inggeris dalam media social.Terdapat 8 fasa dalam aliran penyelidikan iaitu pernyataan masalah. objektif, ulasan sastera, pemahaman rangka kerja, data sentimen memahami, mencadangkan rangka kerja, pengukuran data sentimen dan fasa penilaian keputusan. Ketepatan hasilnya untuk teknik analisis sentimen *(Python NLTK Text Classification, Miopia dan MeaningCloud*) dalam bahasa Inggeris akan dibandingkan. Ketepatan *Python NLTK Text Classification* (pendekatan berasaskan Corpus), *Miopia* (pendekatan berasaskan Lexicon) dan *MeaningCloud* (pendekatan hibrid) adalah 74.5%, 73% dan 82.13%. Ketepatan *MeaningCloud* adalah yang tertinggi di antara 3 teknik analisis sentiment. Ini kerana teknik ini hibrid ciri-ciri pendekatan berasaskan korpus dan berasaskan leksikon dan mencapai ketepatan klasifikasi maksimum. Oleh itu, kaedah tafsiran mesin hibrid dianggap yang terbaik antara tiga model ini.

.

ABSTRACT

Nowadays, numerous numbers of companies have utilized the web to offer their services and products. Web customers dependably look through the comments of other customers towards a product or service before they chose to buy the things or viewed the films. The company needs to analyse their customers' sentiment and feeling based on their comments. The outcome of the sentiment analysis makes the companies easily to discover the expression of their users is more to positive or negative. The sentiment analysis is utilized in data mining. The accuracy of the output is the issue of the sentiment analysis. The objective of this research is to explore and evaluate English language using 3 different sentiment analysis techniques which are the Python NLTK Text Classification, Miopia and MeaningCloud tools in term of their text classification (positive or negative). 3 sentiment analysis techniques have been used in this research to analyse the sentiment analysis of the reviews and comments from English language in social media. There are 8 phases in the research flow which are problem statement. objective, literature reviews, framework understanding, sentiment data understand, propose the framework, measurement the sentiment data and evaluation of the results phases. The accuracy of the output for the sentiment analysis techniques (Python NLTK Text Classification, Miopia and MeaningCloud) in English language will be compared. The exactness of the Python NLTK Text Classification (Corpus-based approach), Miopia (Lexicon-based approach) and MeaningCloud (Hybrid approach) are 74.5%, 73% and 82.13%. The accuracy of MeaningCloud is the highest among the 3 sentiment analysis techniques. This is because this technique hybrids the characteristics of corpus based and lexicon- based approach and achieve the maximum classification accuracy. Therefore, hybrid machine interpretation method is considered the best among these three models.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BNs	Bayesian networks
CMD	Command Prompt
MMH	Maximal margin hyperplane
MPE	Most probable explanation
MT	Machine Translation
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
PCs	Personal Computers
RMBT	Rule Based Machine Translation
SL	Source language
SMT	Statistical Machine Translation
SVM	Support Vector Machine
TL	Target language

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Nowadays, the web has been broadly utilized by the general population universally. Individuals utilize the web to discover films, recreations and books. In the meantime, they additionally express their emotions towards those perspectives with the goal that other people can without much of a stretch to recognize what happens precisely and whether there is any item worth to purchase or attempt through the perception of the surveys. For instance, Twitter has 0.218 billion of dynamic users and 0.50 billion tweets every day(Cardona-Grau, Sorokin, Leinwand, & Welliver, 2016). Subsequently, huge of information has been made and exchange on the web. Big data helps the general population particularly business visionary with the end goal to know whether their items are valuable, and they can know the evaluation from their clients through the perception of the remarks and surveys on the web. The company can carry out sentiment analysis to analyze their customers' sentiment and feeling toward their products and services.

There are different sentiment analysis techniques that exist in the market today to help the companies to conduct the sentiment analysis for their customers' reviews and comments. This help to improve the company's products and services. Sentiment analysis is the information mining that used to investigate the sentiments of the general population towards an element with the end goal to know if individuals like or abhorrence it. It is very troublesome and tedious if enlist a man to dissect the information to know whether the greater part of clients give a positive or negative audit from a huge number of remarks. In this manner, machine learning dependent on sentiment analysis has been presented. It makes the data mining quicker.

1.2 PROBLEM STATEMENT

Table 1.1 Problem Statement

Problem	Problem description	Effect
The sentiment analysis techniques construct their model that only consist of one language.	• The multilingual people can comprehend numerous languages. • They can give the customer reviews in social media using different languages.	It is quite hard for the sentiment analysis technique to extract data. Company cannot interpret their customer reviews from different languages.
Using of words in casual languages and emoji when commenting in social media	• For example, some of the sentiment analysis technique cannot detect the “can’t, wouldn’t “in casual language. • Using the emoji causes some sentiment analysis technique hard to detect the emojis.	This will cause the positive and negative classification in sentiment analysis will be different. This will affect the output of the sentiment analysis.
The accuracy of the yield	• There isn't any sentiment analysis technique that can accomplish the 100% precision. • The different sentiment analysis technique will be used in different situations in order to obtain the better accuracy of sentiment analysis.	Time-consuming in choosing the suitable sentiment analysis techniques.

Source: (Bhuta, Doshi, Doshi, & Narvekar, 2014; Sarlan et al., 2015)

The sentiment analysis techniques construct their model that only consist of one language. The language limitation on its sentiment analysis model makes it hard for multilingual processing. The multilingual people can comprehend numerous languages. They can give the customer reviews in social media using different languages. If the model of sentiment analysis technique only consists of one language, it is quite hard for the sentiment analysis technique to extract the data. For example, Twitter produces enormous amounts of opinion tweets that consist of different languages. It is quite hard for single language model sentiment analysis technique to extract the text and make sentiment analysis. The company cannot interpret their customer reviews from different languages (Sarlan et al., 2015).

Using of words in casual languages and emoji when commenting in social media. For example, some of the sentiment analysis technique cannot detect the "can't, wouldn't "in casual language. While some of the sentiment analysis technique can detect it. This will cause the positive and negative classification in sentiment analysis will be different. Besides that, some of the users tend to use the emoticons to express their expression in social media. Using the emoji to comment is a trend for the media social users. This causes some sentiment analysis technique hard to detect the emojis. This will impact on the evaluation of sentiment analysis technique. This will affect the output of the sentiment analysis. This will cause the positive and negative grouping in sentiment analysis will be different (Sarlan et al., 2015).

The accuracy of the yield is the issue of the sentiment analysis. There isn't any sentiment analysis technique that can accomplish the 100% precision. For instance, I purchased a telephone half a month prior. It is a lovely telephone, although it somewhat enormous in size. In term of sentiment order, it is positive or negative grouping? There are number of sentiment analysis techniques exists in the market nowadays for the users to apply. The different sentiment analysis technique will be used in different situations in order to obtain the better accuracy of sentiment analysis. The general population will ponder which approach should they utilize, and in which circumstance they are required to utilize and bringing about the time-consuming in choosing for the correct approach (Bhuta, Doshi, Doshi, & Narvekar, 2014).

REFERENCES

Al-Aidaroos, K. M., Abu Bakar, A., & Othman, Z. (2010). Naïve Bayes variants in classification learning. *Proceedings - 2010 International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World, CAMP'10*, 276–281. Retrieved from https://ieeexplore.ieee.org/document/5466902

Ang, S. L., Ong, H. C., & Low, H. C. (2016). Classification using the general bayesian network. *Pertanika Journal of Science and Technology*, *24*(1), 205–211. Retrieved from http://pertanika.upm.edu.my/Pertanika PAPERS/JST Vol. 24 (1) Jan. 2016/13 JST-0551-2015 Rev1- Sau Loong Ang.pdf

Avanço, L. V., & Nunes, M. D. G. V. (2014). Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. *Proceedings - 2014 Brazilian Conference on Intelligent Systems, BRACIS 2014*, 277–281. Retrieved from https://ieeexplore.ieee.org/document/6984843

Ballabh, A., & Chandra Jaiswal, U. (2015). a Study of Machine Translation Methods and Their Challenges. *International Journal of Advance Research In Science And Engineering IJARSE*, *8354*(4), 423–429. Retrieved from https://www.ijarse.com/images/fullpdf/320.pdf

Basiri, M. E., & Kabiri, A. (2018). Translation is not enough: Comparing Lexicon-based methods for sentiment analysis in Persian. *18th CSI International Symposium on Computer Science and Software Engineering, CSSE 2017*, *2018–Janua*(Ml), 36–41. Retrieved from https://ieeexplore.ieee.org/document/8320114

Berry, M. W., Mohamed, A. H., & Wah, Y. B. (2015). Reviewing Classification Approaches in Sentiment Analysis. *Communications in Computer and Information Science*, *545*, 43–53. Retrieved from https://link.springer.com/chapter/10.1007/978-981-287-936-3_5

Bhavitha, B. K., Rodrigues, A. P., & Chiplunkar, N. N. (2017). Comparative study of machine learning techniques in sentimental analysis. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017*, (Icicct), 216–221. Retrieved from https://ieeexplore.ieee.org/document/7975191

Bing Liu. (2012). *Sentiment Analysis and Opinion Mining*. Retrieved from https://books.google.com.my/books?id=Gt8g72e6MuEC&printsec=frontcover&dq=Sentiment+Analysis+and+Opinion+Mining&hl=zh-CN&sa=X&ved=0ahUKEwja5fH5xqjeAhUMdCsKHf4CCFYQ6AEIODAB#v=onepage&q=Sentiment Analysis and Opinion Mining&f=false

Bhuta, S., Doshi, U., Doshi, A., & Narvekar, M. (2014). A review of techniques for sentiment analysis of Twitter data. *Proceedings of the 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2014*, 583–591. Retrieved from https://ieeexplore.ieee.org/document/6781346

Cardona-Grau, D., Sorokin, I., Leinwand, G., & Welliver, C. (2016). Introducing the Twitter Impact Factor: An Objective Measure of Urology's Academic Impact on

Twitter. *European Urology Focus*, *2*(4), 412–417. Retrieved from https://www.sciencedirect.com/science/article/pii/S2405456916300086

Copeland, A. (2016). Last lecture summary Naïve Bayes Classifier. Bayes Rule Normalization Constant LikelihoodPrior Posterior Prior and likelihood must be learnt. Retrieved from https://slideplayer.com/slide/7504861/

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y. A., Gelbukh, A., & Zhou, Q. (2016). Erratum to: Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques (Cogn Comput, 10.1007/s12559-016-9415-7). *Cognitive Computation*, *8*(4), 772–775. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981629/

Elkan, T. L. B. (1995). *Unsupervised learning of multiple motifs in biopolymers using expectation maximization*. Retrieved from https://link.springer.com/article/10.1007%2FBF00993379

Fangming Ye, Zhaobo Zhang, Krishnendu Chakrabarty, X. G. (2016). Knowledge-Driven Board-Level Functional Fault Diagnosis, 147. Retrieved from https://books.google.com.my/books?id=QnjhDAAAQBAJ&pg=PA45&dq=strength+and+limitation+of+svm&hl=zh-CN&sa=X&ved=0ahUKEwjjyP3n4OLdAhULSo8KHVMMCXcQ6AEIKzAA#v=onepage&q=strength and limitation of svm&f=false

Foucart, A., & Frenck-Mestre, C. (2016). Natural Language Processing. *The Cambridge Handbook of Second Language Acquisition*, 394–416. Retrieved from https://ieeexplore.ieee.org/document/6678407

Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. *Proceedings - 11th Web Information System and Application Conference, WISA 2014*, 262–265. Retrieved from https://ieeexplore.ieee.org/document/7058024

Ibrahim, D. (2016). An Overview of Soft Computing. *Procedia Computer Science*, *102*(August), 34–38. Retrieved from https://ac.els-cdn.com/S1877050916325467/1-s2.0-S1877050916325467-main.pdf?_tid=0e794e27-aad5-49bd-9ae9-d6597ac31f17&acdnat=1540634369_7a2a8c8f52eebc8e460ae2a7097b799a

Jadav, B., & Vaghel, V. (2016). Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis. *International Journal of Computer Applications*, *146*(13), 26–30. Retrieved from https://pdfs.semanticscholar.org/7746/93175a159160697b5748561446313186b846.pdf

Kazemi, A., Toral, A., Way, A., Monadjemi, A., & Nematbakhsh, M. (2017). Syntax-and semantic-based reordering in hierarchical phrase-based statistical machine translation. *Expert Systems with Applications*, *84*, 186–199. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417417303160

Kumpikova, N., & Kumpikova, N. (2018). Hybrid Machine Translation Architectures : The EuroMatrix Project. Retrieved from http://www.sfs.unituebingen.de/~keberle/HybridMT/Presentations/preziACS.pdf

LANGE, W. (2017). The Pros and Cons of Statistical Machine Translation. Retrieved October 25, 2018, from http://daily.unitedlanguagegroup.com/stories/pros-cons-statistical-machine-translation

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2017). Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review*, *48*(4), 499–527. Retrieved from https://link.springer.com/content/pdf/10.1007%2Fs10462-016-9508-4.pdf

Loon, R. van. (2018). Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning. Retrieved from http://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/

Maglogiannis, I. G. (2007). *Emerging Artificial Intelligence Applications in Computer Engineering*. Retrieved from https://books.google.com.my/books?hl=zh-CN&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=Supervised+machine+Learning&ots=CYnyzs2Djo&sig=U0SlP0BY7JRIyHM2BobVrkqvOVI&redir_esc=y#v=onepage&q=Supervised machine Learning&f=false

MeaningCloud. (2018). What is a model? Retrieved from https://www.meaningcloud.com/developer/text-classification/doc/1.1/what-is-model?fbclid=IwAR3mqtT6ZwD7ULIGkbdqnpm7qTe3BYdufVJPse5LGBNnFOS3ZXmgeSaE8RU

Minn, S., Fu, S., & Desmarais, M. C. (2014). Efficient learning of general Bayesian network classifier by local and adaptive search. *DSAA 2014 - Proceedings of the 2014 IEEE International Conference on Data Science and Advanced Analytics*, (2), 385–391. Retrieved from https://ieeexplore.ieee.org/document/7058101

Moussallem, D., Wauer, M., & Ngomo, A. C. N. (2018). Machine Translation using Semantic Web Technologies: A Survey. *Journal of Web Semantics*, *51*, 1–19. Retrieved from https://www.sciencedirect.com/science/article/pii/S1570826818300301

Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*. Retrieved from https://ieeexplore.ieee.org/document/6726818

Nijhawan, R., Roorkee, T., Srivastava, I., & Shukla, P. (2017). Land Cover Classification Using Supervised and Unsupervised Learning Techniques, 1–6. Retrieved from https://ieeexplore.ieee.org/document/8272630

Ochieng, S. B. O., Loki, M., & Sambuli, N. (2016). Limitations of sentiment analysis on Facebook data. *International Journal of Science and Information Technology*, (June), 425–433. Retrieved from https://www.researchgate.net/publication/304024274_LIMITATIONS_OF_SENTIMENT_ANALYSIS_ON_FACEBOOK_DATA

Okpor, M. D. (2014). Machine Translation Approaches : Issues and Challenges, *11*(5), 159–165. Retrieved from https://www.ijcsi.org/papers/IJCSI-11-5-2-159-165.pdf

Patil, P., & Yalagi, P. (2016). Survey on Aspect-Level Sentiment Analysis.pdf, *6*. Retrieved from http://ijiet.com/wp-content/uploads/2016/05/72.pdf

Petr Sojka, H. S. et al. (2018). Introduction to Information Retrieval. Retrieved from https://www.fi.muni.cz/~sojka/PV211/p15svm.pdf

Pong-inwong, C., & Songpan, W. (2018). Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining. *International Journal of Machine Learning and Cybernetics*, *0*(0), 0. Retrieved from http://link.springer.com/10.1007/s13042-018-0800-2

Professor, M. S., & Technology, L. & I. (2018). Learning, explanations of sentiment analysis with supervised. Retrieved from https://www.coursera.org/lecture/text-mining-analytics/5-1-explanations-of-sentiment-analysis-with-supervised-learning-hbTb7

Rabab'Ah, A. M., Al-Ayyoub, M., Jararweh, Y., & Al-Kabi, M. N. (2016). Evaluating SentiStrength for Arabic Sentiment Analysis. *Proceedings - CSIT 2016: 2016 7th International Conference on Computer Science and Information Technology*. Retrieved from https://ieeexplore.ieee.org/document/7549458

Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-Based Sentiment Analysis of Teachers' Evaluation. *Applied Computational Intelligence and Soft Computing*, *2016*, 1–12. Retrieved from https://www.hindawi.com/journals/acisc/2016/2385429/

Rouse, M. (2010). artificial neural network (ANN). Retrieved from https://searchenterpriseai.techtarget.com/definition/neural-network

Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). SentiCircles for contextual and conceptual semantic sentiment analysis of Twitter. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8465 LNCS*, 83–98. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-07443-6_7

Samariya, D. (2016). *Tweelyzer. An Approach to Sentiment Analysis of Tweets*. Retrieved from https://books.google.com.my/books?id=Mn43DQAAQBAJ&pg=SA5-PA17&dq=SENTIMENT+ANALYSIS+An+Overview.+Comprehensive+Exam+Paper+MEJOVA&hl=zh-CN&sa=X&ved=0ahUKEwiGi7HJwKjeAhUhT48KHbMRDXcQ6AEILDAA#v=onepage&q=SENTIMENT ANALYSIS&f=false

Sarlan, A., Nadam, C., & Basri, S. (2015). Twitter sentiment analysis. *Conference Proceedings - 6th International Conference on Information Technology and Multimedia at UNITEN: Cultivating Creativity and Enabling Technology Through the Internet of Things, ICIMU 2014*, 212–216. Retrieved from https://ieeexplore.ieee.org/document/7066632

Sha, Y., & Liu, J. (2013). Sentence semantic orientation calculation algorithm based on structure analysis. *Proceedings - 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation, IMSNA 2013*, 110–113. Retrieved from https://ieeexplore.ieee.org/document/6742828

Shravan, I.V. (2016). Analysing Sentiments with NLTK. Retrieved from https://opensourceforu.com/2016/12/analysing-sentiments-nltk/

Tripathy, P., Rautaray, S. S., & Pandey, M. (2017). Role of Parallel Support Vector Machine and Map- reduce in Risk analysis, 9–11. Retrieved from https://ieeexplore.ieee.org/document/8117736

Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing and Management*, *53*(3), 595–607. Retrieved from http://sci-hub.tw/https://doi.org/10.1016/j.ipm.2017.01.004

Vilares, D., Gómez-Rodríguez, C., & Alonso, M. A. (2017). Universal, unsupervised (rule-based), uncovered sentiment analysis. *Knowledge-Based Systems*, *118*, 45–55. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0950705116304701

Xuan, H. W., Li, W., & Tang, G. Y. (2012). An advanced review of hybrid machine translation (HMT). *Procedia Engineering*, *29*, 3017–3022. Retrieved from https://ac.els-cdn.com/S1877705812004420/1-s2.0-S1877705812004420-main.pdf?_tid=ec681d38-4168-45e3-9ae2-01978c74af70&acdnat=1540794046_82fe70f9f43901db79218b7de3900fff

Zhang, H., & Li, D. (2007). Naive Bayes text classifier. *Proceedings - 2007 IEEE International Conference on Granular Computing, GrC 2007*, (3), 708–711. Retrieved from https://ieeexplore.ieee.org/abstract/document/4403192