

# Machine learning models to predict psychometric personality types in order to better understand customer reviews

Christopher Culley  
University of Southampton  
Southampton, Hampshire, UK  
cc2u18@soton.ac.uk

Sam Banks  
University of Southampton  
Southampton, Hampshire, UK  
swb1n18@soton.ac.uk

Fairouz Aldabbagh  
University of Southampton  
Southampton, Hampshire, UK  
fa2n18@soton.ac.uk

Jak Hall  
University of Southampton  
Southampton, Hampshire, UK  
jh16g18@soton.ac.uk

Claudia Subia  
University of Southampton  
Southampton, Hampshire, UK  
cms2n17@soton.ac.uk

## ABSTRACT

In this work we present a predictive model pipeline for determining the personality of business users utilising comments supporting business review feedback. We use forum comments with attached Myers Briggs personality types in building the models with high accuracy, a mean of 85% for each of the four letters, and use these models to determine the personality types of Yelp business reviews. We demonstrate the utility of such a pipeline in a web application which allows businesses to understand its reviewer base across the personality types.

## 1 INTRODUCTION

Understanding business reviews can lead to positive structural reforms and highlight areas of progress. Moderate reviews with depth have been shown to be the most beneficial in gaining a deeper understanding of a customer base [1] and extreme reviews can have a detrimental effect on service use [2]. Exploring these two, and other, sets of reviewer personality traits could therefore be a boon to businesses in identifying and responding to customer trends.

Previous studies have explored the personology of consumer behaviour [3] and the role of personality traits in online transactional decision making [4] and also personality detection from a corpus of words [5] but an emphasis on identifying personality traits in determining business reviewer ratings distributions, to the best of our knowledge, is novel.

In this work we present an analytic application which could help business owners determine which personality types use their service and how favourably, or otherwise, those personality types reviewed the experience.

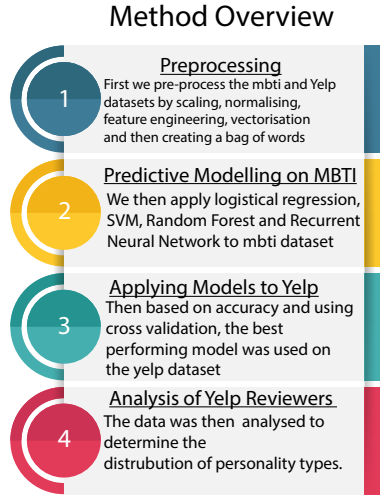
To do so we use a combination of text-processing, machine learning and statistical analysis in determining the personality types of business reviewers and the distribution of ratings across selected businesses which we present in a web application.

To understand the personality of users the MBTI (Myers-Briggs Type Indicator) [6] can be used to categorise users into personality

types. The type indicator is made up of four key factors: Extroversion/Introversion, Sensing/Intuition, Thinking/Feeling and Judging/Perceiving which, together, make up a personality type.

We next give a brief summary of each indicator but direct the reader to [6] for more details. Extroverted (E) people are energised by social interaction and prefer to share problem solving. Introverted (I) people are drained by social interaction, preferring quiet reflection in their own inner world. They prefer one or two key friendships to group settings. Sensing (S) refers to people who are concerned with practical physical reality and acting in the moment, whilst Intuition (N) refers to people who are less practical, preferring to think things through. A Thinking (T) person processes information based on logic and reason whilst a Feeling (F) person accounts for other people's feelings, tending to make decisions with their hearts. A Judging (J) person likes to feel in control and lives an ordered, well organised life. A Perceiving (P) person is more flexible and open to new experiences, without planning ahead. A four-letter system is applied to personality types so, for example, an extroverted, sensing, thinking and judging person would be labeled ESTJ.

We present the overview pipeline in Figure 1; we start first with pre-processing the text datasets ready for the machine learning algorithms utilising extensive feature engineering for added predictability, we then build separate models using logistical regression, support vector machines, random forest and a feed-forward neural network applied to a dataset which has user comments linked to their personality type. After comparing the accuracy of each of the models, we use the best four models (one for each binary choice of personality letter) and apply them to a business review dataset linking business reviews with personality types. The final step incorporates these statistical results into a web application. This paper is structured as follows, we first give details of the datasets used to build the predictive model and then applied on to demonstrate the application. Next we describe the pre-processing steps applied to the raw data followed by the machine learning pipeline. After this, the results section gives the predictive modeling which is followed by the application overview. We close with a discussion and conclusion.



**Figure 1: An overview of the method used to create the application, showing each stage in the developmental process in order of occurrence.**

## 2 DATA

We use the Yelp dataset [7] as data for which we aim to understand the distribution of personality of reviewers. This dataset has been used to build predictive models of business ratings through user text alone [8] and extracting customer subtopics [9] owing to its large size and data integrity.

This data supplies over 5 million reviews of all types of businesses for eleven cities across four countries. For our purposes, and for computational limitations, we focus on Las Vegas (USA) but note to the reader that scaling the prototype application to cover the entire dataset is possible with hardware expansion.

In order to categorise the reviewers into their personality types we first build a predictive model using and Mbt comments dataset [10] taken from the PersonalityCafe forum. This dataset contains 8600 users each with multiple comments on the forum accompanied by the users self-proclaimed personality type.

## 3 ETHICS

In order to comply with data protection regulations, we note that people do not use ‘handles’ or usernames when leaving comments, instead proper names are used. These names are removed in order to preserve anonymity, even though only first names are used in most cases. Secondly, although the comments are used to process and classify personality types, no comment is used in the output data, thus preserving total anonymity. This also prevents the possibility of anyone seeing their comment and taking offense at being classified as a personality type which did not reflect their self-image.

## 4 PREPROCESSING

Biased datasets reduce the generalisability of predictive models and as such need to be corrected. The MBTi, has clear bias with some letter weightings observed for 75% of the datapoints. Although there are many options to resolve this [11], we utilise the simplest though re-sampling separate balanced datasets by target. We do this for each of the four targets in the MBTi dataset, creating four distinct train test sets with balanced targets.

In order to ready the data to build the predictive models we first apply a number of pre-processing steps in order to maximise the opportunity for meaningful patterns to be extracted. We begin by combining reviews from the same person in both the datasets into corpus from which we apply some feature engineering listed in Table 1. We note to the reader, the sentiment analysis features, polarity and subjectivity as well as the word probability information were applied, using the NLTK library [12], to each users with the listed statistics captured over the individuals aggregated corpus.

We next turn the corpus into a bag-of-words where each instance of a word in a corpus corresponds to count in a column. We do this for the 5,000 most common lemmatised words found inside the MBTi dataset ignoring stopwords. We take forward in total 5,017 features to be used inside the machine learning models.

At this point we note to the reader the quandary of scaling. Since scaling the data is a requirement for some algorithms, such as the SVM, and leads to faster solution convergence for neural networks [13] we need to apply scaling to the MBTi. In situations when it is expected that train-test splits are of the same distributions one would normally minus the train mean and divide by the train standard deviation for all data-points in both the train and test set which, in the MBTi case, presents no problem.

The problem arises when we want to next apply the model learned from the MBTi to the Yelp data. Which, may follow a different distribution, this is particularly pronounced when there is more data per corpus (and thus a higher average word count). To overcome this, we normalise the Yelp data using its own mean and standard deviation which we found empirically gave a distribution of targets closer to which is expected in the general population but note this highlights a limitation in our generalisation approach – applying machine learning models from one source to another.

## 5 MACHINE LEARNING

To build the predictive models which we will later apply to the Yelp dataset, we explore four machine learning models which we apply to each of the binary decisions available that construct together to create the four letter personality type. The machine learning algorithms we explore are as follows: logistical regression – selected for its simplicity as a benchmark, support vector machines – shown to work well with prediction models for text applications, random forests – an ensemble method popular also for text prediction problems and a feed-forward neural network – for its ability to construct highly complex latent variables.

Features Extracted
Average word length
Max polarity
Min polarity
Average polarity
Max subjectivity
Min subjectivity
Average subjectivity
Percentage of words misspelled
Average misspelled word length
Max word probability
Average word probability
Standard deviation of word probabilities
Percentage of words that are emoticons
Percentage of letters that are punctuation
Percentage of words that are uppercase (excluding single letters)
Percentage of letters that are numerical
Percentage of words that are stop words

**Table 1: The features extracted from the corpus of words defined by each user in both the MBTi and the Yelp datasets**

For each of the predictive model training we use a pipeline of 5 fold cross validation to tune any hyper-parameters and calculate the accuracy scores using a held out test set. We next discuss briefly each of the algorithms.

### 5.1 Logistic Regression

Logistic Regression is a regression algorithm commonly used for classification. It uses a logistic function to calculate the probability of the data belonging to a class or not and outputs it as a binary value between 0/No and 1/Yes [14]. Logistic Regression allows working with highly interdependent features and it can support millions of features in the model and there is no need to define similarity function. It is faster at classification other standard methods such as k nearest-neighbors. This method has few parameters, but since we use gradient decent as an optimisation procedure we determine before hand the maximum number of iterations.

### 5.2 SVM

The support Vector Machine (SVM) is a supervised machine learning algorithm used for both linear or nonlinear classification, regression (SVR variant), and outlier detection[15]. An SVM is able to project the data points into some new vector space of potentially infinite dimensions where the data is linearly separable. Importantly, rather than have to project all data points into this space, the algorithm only requires a similarity calculation between data points in this new space making the SVM computationally efficient for even the most complex projections.

The SVM is well suited to complex classification problems with small or medium-sized datasets. SVM models are used in text classification problems due to strong performance when working with

high dimensional data such as text [16, 17].

In order to use this model it is important to scale the data because SVM models are very sensitive to feature scales [18] which we account for in the pre-processing stage.

We use an radial basis function kernel to project the data and tune the error allowance of the Lagrangian multipliers in the cross validation stage.

### 5.3 Random Forest

Random Forest [19] are ensemble classification models based on multiple decision trees, which use entropy to select conditional splits leading to prediction outputs, trained using different combinations of the available data and features with bagged voting outputs.

Random forests benefit from having very few hyper parameters to tune such as the minimum tree height and the minimum samples in a split which we again find using cross validation.

### 5.4 Feed Forward Neural Network

A feed forward neural network is an acyclic network of nodes with weighted connections between distinct layers with non-linear activation functions applied to the summed input of a node to give the output which is passed as input to the next layer [20].

After grid searching the hyper parameters we opt for the input vectors connected to four hidden layers each with 75 neurons. We follow others in using dropout and maxnorm regularisation [21] to reduce over-fitting. We use Relu activation function on the inner layers and a binary cross-entropy on the output with ADAM optimiser [22].

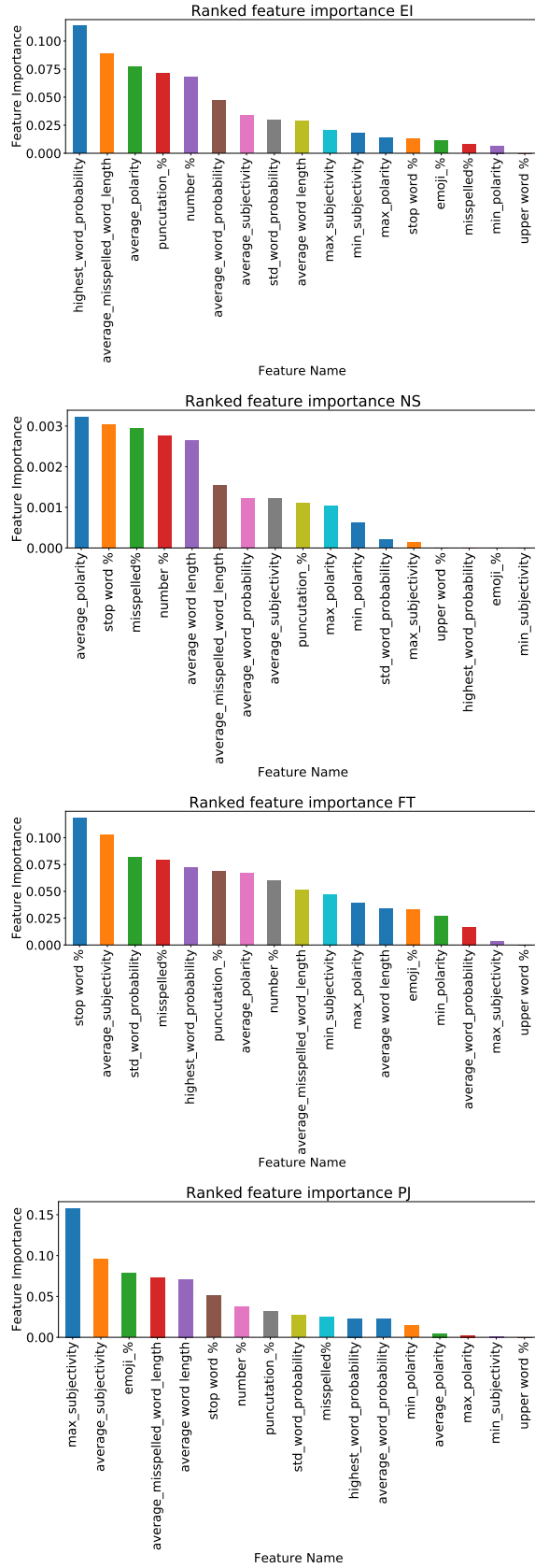
## 6 MACHINE LEARNING RESULTS

We present the machine learning results in Table 2 and schematically in Figure 3 . We achieve strong accuracy rates across each of the models with high accuracy rate. Taking the strongest performing model for each of the classification problems we score an mean accuracy of 86%. We note that the strongest performing model was logistical regression in all but one case which, as one of the simplest methods available, demonstrates the importance of feature engineering in building text classification models.

In Figure 2 we give the engineered features ranked by importance for each of the models, measured by contributing coefficient. We note with interest the large variability of ranking of these features depending on the letter being predicted.

## 7 ANALYSIS

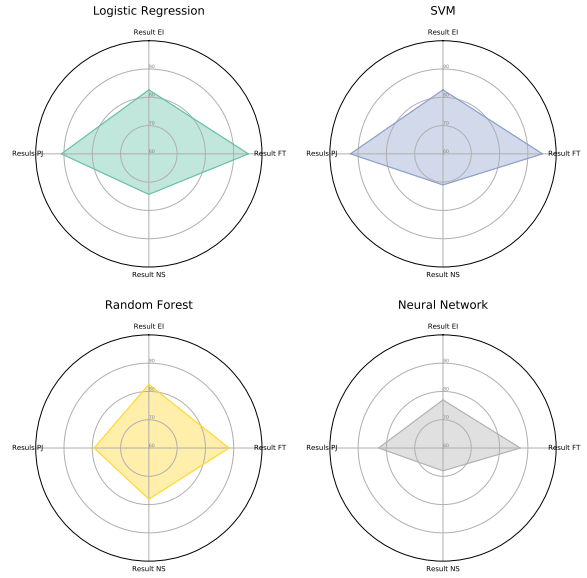
After applying the best performing predictive models as selected by accuracy against the test set in the previous section to the Yelp dataset, we obtain a four letter personality for each of the reviewers. We present next some key findings.



**Figure 2: The ranking of engineered features in each of the predictive models.**

ML Models Results				
ML Model	E/I	F/T	N/S	P/J
Logistic Regression	0.826	0.952	0.743	0.909
SVM	0.826	0.952	0.709	0.927
Random Forest	0.825	0.883	0.780	0.793
Neural Network	0.770	0.873	0.680	0.827

**Table 2: The accuracy scores of the best predictive models using each of the machine learning approaches applied to each of unseen test sets for each of the letters that constitute a personality type according to Myres Briggs typology.**



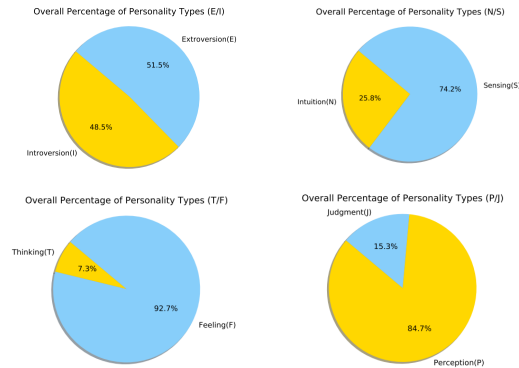
**Figure 3: Here we show the accuracy scores for each of the machine learning approaches in building predictive models for the 4 binary letter choices.**

Firstly, the distribution of personalities in Figure 4. In determining the validity of these results we look to research into the true distribution of personality types, although there are some samples collected for teachers and project managers [23–25], as far as we are aware, there are no meta studies of the general population. We instead look to the MBTI manual [26], which estimates the global distribution from multiple of sources and present the most frequent types in Table 3. Here we can see that whilst the E/I and N/S letters follow a distribution similar distribution of that found in the general population, that there is some discrepancy between the latter two letters F and J. It is difficult to know for certain whether the true distribution of personality letters in the Yelp dataset are closer to that of our results and of that found in [27] but it is a point of notice which warrants further investigation.

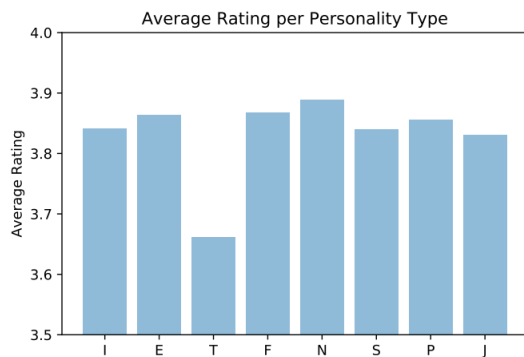
Machine learning models to predict psychometric personality types in order to better understand customer reviews

Letter	Yelp distribution	MBTI distribution	Difference
I	48.5%	42%	+6.5%
N	25.8%	26.6%	-0.8%
F	92%	59%	+37%
J	15.3	54.5%	-39%

**Table 3: A table comparing the distribution of personality letters in the Yelp dataset as determined by the predictive model and that found in the MBTI manual [26]**

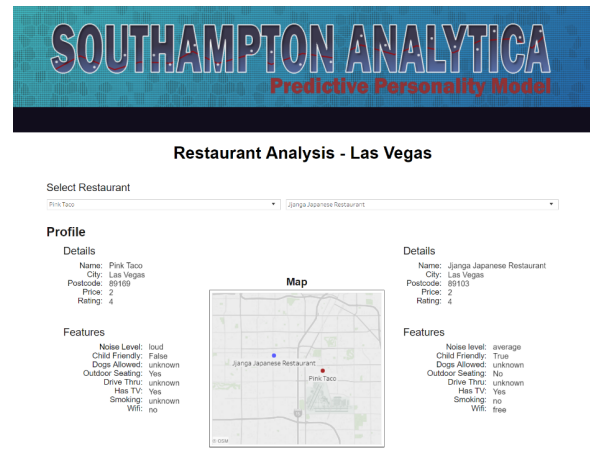


**Figure 4: The distribution of personalities found after applying the predictive model to the Yelp dataset.**

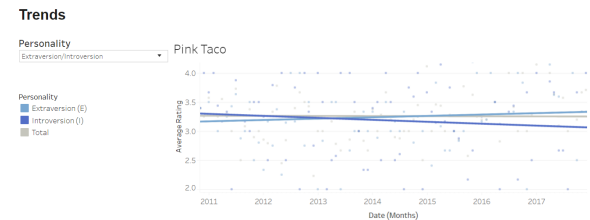


**Figure 5: The average ratings given in the Yelp dataset by personality type**

Interesting, we see a small but notable difference in the distribution of ratings given by each of the letter types, which we give in Figure 5. Whilst, this difference is small, we do notice a larger variance of ratings by letter depending on which business the reviews are for. It is this, the difference in ratings allocated by personality by business which motivates the application which we next discuss.



**Figure 6: The main application view, where businesses can be selected and mapped based on locational data**



**Figure 7: The trends chart in the application. In this example we observe that the personality average rating switches in 2013, despite the total average rating staying relatively static. This infers that the business has appealed more to extroverted customers in recent years.**

## 8 APPLICATION

We next present a personality model application. For demonstrative purposes, we take the results of the Yelp data with appended personalities after applying the machine learning models and produce an application to inform Yelp using business.

The application is intended for broad analysis, allowing for two businesses in the processed dataset to be compared against each other. It displays relevant metrics already present in the Yelp dataset, such as price ranges and location (shown in Figure 6) and provides a word cloud, generated from the 50 most frequent terms (after eliminating stop words) giving an overview of the language being used.

Additionally it allows the business to explore how ratings are distributed by personality types. We achieve this by graphing the distribution of user ratings for each of the applicable personality types as well through trend charts, which reflect the average user

ratings overtime by personality (presented in Figure 7).

This added functionality enables the tool to be used by a business to evaluate and understand trends in ratings given by reviews of varying personalities.

We further motivate the application with the following: If a business wants to improve its customer satisfaction, as captured in rating response, should it concentrate on moderate service users that are of a personality type that it currently enjoys success with? Or should it look to similar businesses which are doing well with customers of a personality type that it is not, and try to understand and emulate its success?

Our analysis and application allows for such questions to be asked where before they could not. Further iterations will look to build upon these features with the intention to gain feedback from potential users of the business of any improvements.

## 9 DISCUSSION

We start with some limitations with our works. Firstly the use of MBTI as a way of categorising personality has some notable limitations [27] and another set of personality determinants, such as the Big Five [28] could very well be more informative and considered for future models.

We also note that building predictive models on one source of data and using this model to make predictions on another source of data can prove inaccurate since the underlying distributions, and therefore patterns used in making the predictions, could be very different leading to inaccuracies. This could account for the variation in observed distributions of personalities in the Yelp data and of that found in the literature.

That said, our models have strong predictive accuracy which are able to, given some strong assumptions about the universal nature of language structure of persons of particular personality types, be applied to other similar text datasets to create a prototype application of commercial use.

We envision that, with further work improving the models ability to be applied to multiple sources, this application could be used in understanding the distribution of personalities in a wider set of contexts.

## 10 CONCLUSION

We have explored an application of machine learning in determining the personality of Yelp business reviews. Using these predictions we then demonstrated the potential utility through the building of a review analytics application which provides a granular understanding of the role of personalities in the distribution of review ratings. Our results show that the distribution of ratings vary across different personality types and that these change over time. This could in the future allow businesses to monitor and react to user

base responses to marketing strategies at a personality trait level and give valuable insight into the nature of reviewers.

## 11 APPLICATION LINK

Available at: <https://jakhall.github.io/southampton-analytica/>

Email login: [sa.guest@email.com](mailto:sa.guest@email.com)

Password: guest.19

## REFERENCES

- [1] S. M. Mudambi and D. Schuff, "Research note: What makes a helpful online review? a study of customer reviews on amazon. com," *MIS quarterly*, pp. 185–200, 2010.
- [2] B. A. Sparks and V. Browning, "The impact of online reviews on hotel booking intentions and perception of trust," *Tourism management*, vol. 32, no. 6, pp. 1310–1323, 2011.
- [3] H. Baumgartner, "Toward a personology of the consumer," *Journal of consumer research*, vol. 29, no. 2, pp. 286–292, 2002.
- [4] M. Bosnjak, M. Galesic, and T. Tuten, "Personality determinants of online shopping: Explaining online purchase intentions using a hierarchical approach," *Journal of Business Research*, vol. 60, no. 6, pp. 597–605, 2007.
- [5] F. Mairesse and M. Walker, "Automatic recognition of personality in conversation," in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pp. 85–88, Association for Computational Linguistics, 2006.
- [6] M. H. McCauley, "The myers-briggs type indicator: A measure for individuals and groups," *Measurement and evaluation in counseling and development*, vol. 22, no. 4, pp. 181–195, 1990.
- [7] Y. D. Challenge, "Yelp dataset challenge," 2013.
- [8] M. Fan and M. Khademi, "Predicting a business star in yelp from its reviews text alone," *arXiv preprint arXiv:1401.0864*, 2014.
- [9] J. Huang, S. Rogers, and E. Joo, "Improving restaurants by extracting subtopics from yelp reviews," *iConference 2014 (Social Media Expo)*, 2014.
- [10] PersonalityCafe, "Mbt-type - personalitycafe dataset," 2016.
- [11] J. Huang, A. Gretton, K. M. Borgwardt, B. Schölkopf, and A. J. Smola, "Correcting sample selection bias by unlabeled data," in *Advances in neural information processing systems*, pp. 601–608, 2007.
- [12] S. Bird and E. Loper, "Nltk: the natural language toolkit," in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, p. 31, Association for Computational Linguistics, 2004.
- [13] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*, pp. 9–48, Springer, 2012.
- [14] D. Hosmer and S. Lemeshow, "Applied logistic regression., 2nd edn.(wiley: New york.)," NY, USA, 2000.
- [15] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2017.
- [16] Z.-Q. Wang, X. Sun, D.-X. Zhang, and X. Li, "An optimal svm-based text classification algorithm," in *Machine Learning and Cybernetics, 2006 International Conference on*, pp. 1378–1381, IEEE, 2006.
- [17] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [18] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, pp. 137–142, Springer, 1998.
- [19] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [20] R. Lippmann, "An introduction to computing with neural nets," *IEEE Assp magazine*, vol. 4, no. 2, pp. 4–22, 1987.
- [21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [23] Y. Cohen, H. Ornoy, and B. Keren, "Mbt personality types of project managers and their success: A field survey," *Project Management Journal*, vol. 44, no. 3, pp. 78–87, 2013.
- [24] N. Levy, C. Murphy Jr, and R. Carlson, "Personality types among negro college students," *Educational and Psychological Measurement*, vol. 32, no. 3, pp. 641–653, 1972.
- [25] A. Furnham and P. Stringfield, "Personality and occupational behavior: Myers-briggs type indicator correlates of managerial practices in two cultures," *Human Relations*, vol. 46, no. 7, pp. 827–848, 1993.
- [26] I. B. Myers, M. H. McCauley, N. L. Quenk, and A. L. Hammer, *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*, vol. 3. Consulting Psychologists Press Palo Alto, CA, 1998.

## Machine learning models to predict psychometric personality types in order to better understand customer reviews

- [27] G. J. Boyle, "Myers-briggs type indicator (mbti): Some psychometric limitations," *Australian Psychologist*, vol. 30, no. 1, pp. 71–74, 1995.
- [28] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr, "A very brief measure of the big-five personality domains," *Journal of Research in personality*, vol. 37, no. 6, pp. 504–528, 2003.