COMP6246 Machine Learning Tech Report

Zhe Shen
Student number: 31533914
Email: zs3m20@soton.ac.uk

1 Introduction and Data Analysis

1.1 Introduction

The popularity of social media makes it easier and faster for people to get information in their daily life. However, it also brings many problems, the most serious of which is the spread of fake information. It also caused a lot of trouble for people. The main purpose of this research is to use machine learning techniques to automatically detect the authenticity of tweets and prompt users on time. In this way, people are less likely to be misled when they acquire information on social media.

Various techniques can complete this task [1]. In the research of Lim, they achieve good performance by using KNN to classify text [2]. Beyond that, Support vector machines (SVMs) are also a popular class of machine learning algorithms used for text classification [3]. In the area of traditional machine learning algorithms, Naïve Bayes is also widely used in the task of classification considering its simplicity and effectiveness [4]

With the popularity of neural networks, many researchers applied neural networks in the task of text classification in recent years. Harrag proposed a method that applied Artificial Neural Network in the classification of Arabic language documents and achieved 0.88 accuracies with the ANN model using SVD [5]. As an effective algorithm in image classification, Convolutional Neural Network trained on top of pre-trained word vectors is also applied in the field of sentence classification [6].

Another aim is to compare the above algorithms. Which algorithm can give the best performance? Which algorithm can run fastest? What are the strengths and weaknesses of these methods? Before giving answers, we first introduce our datasets and how to preprocess them.

1.2 Data Analysis

Data Format

Our dataset is from the Mediaeval 2015 which can download from the ECS student intranet.The Mediaeval 2015 contained three parts, the main parts that we are going to use are medieval-2015-training set and the mediaeval-2015-test set. Both contained tweetId, tweetText, userId, imageId, username, timestamp, and label. The format of these two datasets is the text encoded by UTF-8.

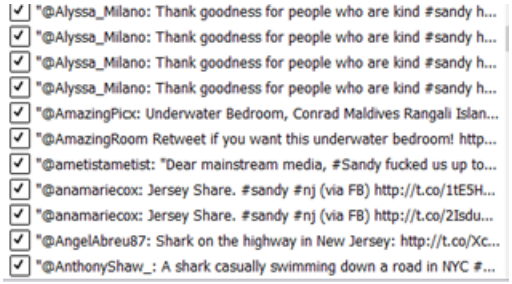

Figure 1: Retweet Text.

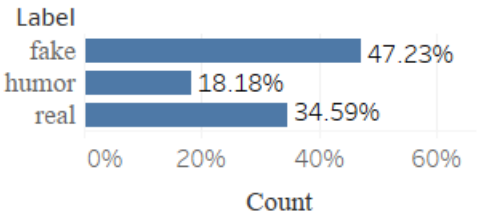

Figure 2: Histogram of training set labels.

Data Number

The training set contained 14483 tweets and the test set contained 3781 tweets.

Data Quality

The key operation to analyzing the quality of data is to understand every field. In all fields, tweet text represents key information of these data. But some of them are incomplete or contain repeated information, for instance, a huge number of tweet texts have repeated @ other users' tweets as shown in the figure 1. Secondly, many tweet texts contain redundant information like images websites, and emoji characters, this redundant informa-

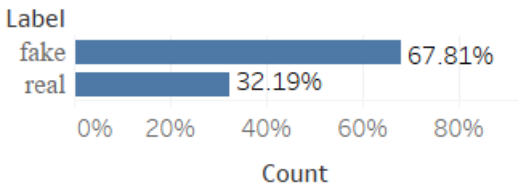

Figure 3: Histogram of testing set labels.

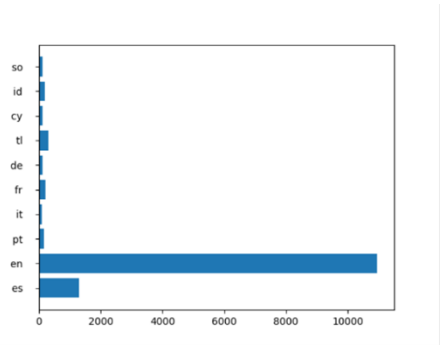

Figure 4: Top 10 languages with the largest number of training sets.

tion is hard to transform in the training process. Then some tweet text contains mixed language and some of these texts even have bad grammar problems.

Another key field is the label, our task is to classify a tweet as fake or real, but there contained humor in the label column.

Data Bias Legal and Ethics

There are 6841 fake label data and 2633 humor label data while there are only 5009 real label data in the training dataset.

In the testing data, there are 2564 fake label data and there are only 1217 train label data. Therefore, the data set is unbalanced. In terms of language bias, we found that tweets text included 44 kinds of language in the training set. Figure 4 shows the share of the top 10 languages in the training set.

In the testing set, there are 34 kinds of language, different kinds of language will impact classification accuracy.

The above problems will be dealt with in the data preprocessing stage. Our datasets are open to students and are completely legal to use in legal and moral terms.

2 Data Preprocess and Algorithm Design

2.1 Data Preprocess

For the above data quality problems, the first thing this research needs to do is data cleaning. This research would choose tweet text and label as key fields to process.

Remove repeated retweets

Considering there are a lot of repeated retweets in the data, we removed those retweets and repost tweets in case of data duplication. Firstly, we defined some patterns by using regular expressions to represent retweeting data. Then we removed these data that contained the above patterns by using pandas which is a Python library.

Remove emojis

Parts of tweets included some emojis that are hard to identify, one solution to solve this problem is to remove these emojis.

Remove URL

Another problem is most tweets contain web addresses, which may also affect text recognition, therefore, this paper reduces them by doing the same operation as removing retweets.

Remove newline character

Also, some data includes newline characters which could influence the quality of data, this paper also remove these new line characters.

lowercase

In this research, all the words in the dataset were converted to lowercase to avoid the model classifying 'world' and 'WORLD' as two different classes.

Stop words

In the task of text classification, data quality can be effectively improved by removing stop words. The data would be filtered by comparing with a list of commonly used stop words in this research.

The quality of data has improved significantly after the above preprocessing. However the problem of data bias still exists, As for language bias, this research translated all non-English languages into English. As for label bias, considering in real networks, there are more fake messages than real messages [7], so the bias of labels will not processed except converting label humor to fake.

Translate tweet text

This research translated all the tweets into English by calling Google Trans library in Python. We can mitigate language bias through this method.

Tokenization

Tokenization splits the text into a sequence of strings (elements commonly called tokens, or words) according to specific requirements. This operation aims to better analyze the content of the text information and the meaning the text wants to express.

Some tweets in the data set that fail to be translated or forwarded will be deleted. Therefore, there are 11546 training data and 3233 testing data in the datasets now. Parts of processed data are shown in the figure5.

```
Found the missing Malaysian plane !
Facebook malware : Malaysian MH370 plane found in
Courtney Love : I may have found missing Malaysian
Malaysian plane finally found !
```

Figure 5: Parts of processed data.

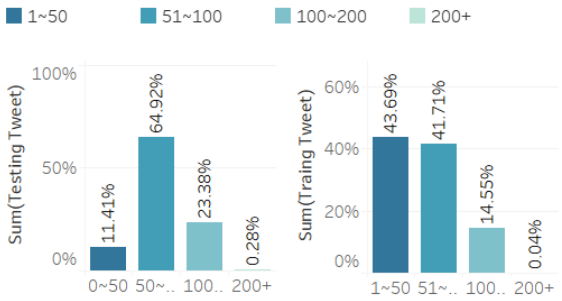

Figure 6: The data length of dataset.

2.2 Algorithm Design

2.2.1 K-Nearest Neighbor

K-Nearest Neighbor is one of the simplest methods in data mining classification technology. The reason that chooses KNN in this research is Khamar used KNN to get better accuracy than other algorithms in the task of short text classification [8]. Most of the tweets that were processed are short texts, the data length of the dataset is shown in figure 6. Therefore, KNN is suitable for this task.

2.2.1.1 Feature Selection

In information retrieval, the Bag of Words model assumes that for a text, its word order, grammar, and syntax are ignored and it is only regarded as a word set or a combination of words. The occurrence of each word in the text is independent of the occurrence of other words.

Implement CountVectorizer from sklearn to implement Bag of words model in this task. Every tweet text will transform into a 1*8265 text vector as shown in figure7.

2.2.1.2 Dimensionality Reduction

There contain more than 1000 features for one text after using the bag of words model to extract features, which could be the curse of dimensionality(Too many dimensions may degrade the accuracy of the classifier).

```
shape: (11544, 8265)
Do you remember the movie : " The day after tomorrow " ? It reminds me of what is happening with Hurricane Sandy .
[0 0 0 ... 0 0 0]
Good Hurricane Sandy ' s photo reminds me of the independence day movie id4 sandy
[0 0 0 ... 0 0 0]
Scary shit hurricane NY
[0 0 0 ... 0 0 0]
My fave place in the world nyc hurricane sandy statueofliberty
[0 0 0 ... 0 0 0]
```

Figure 7: Selected 1*8265 vector.

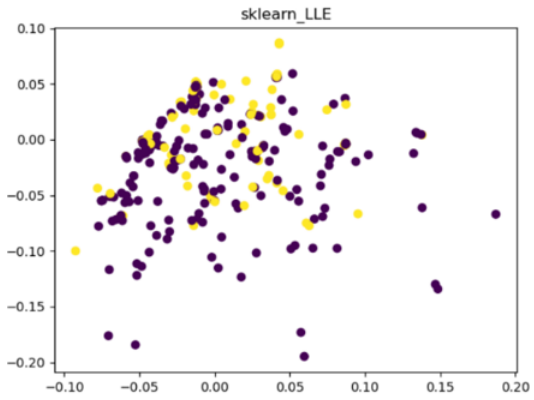

Figure 8: The scatter diagram of partial random data processed by dimensionality reduction to two dimensions.

Introduce Locally linear embedding(LLE) to solve this problem, the LLE algorithm considers that each data point can be constructed from a linear weighted combination of its neighboring points [9]. The main steps of the algorithm are divided into three steps :(1) to find k nearest neighbor points of each sample point; (2) The local reconstruction weight matrix of each sample point is calculated by the nearest neighbor of each sample point; (3) The output value of the sample point is calculated from the local reconstruction weight matrix of the sample point and its nearest neighbor points.

The figure8 is a scatter diagram of partial random data processed by dimensionality reduction to two dimensions.

2.2.1.3 Algorithm Introduction

K nearest neighbors mean that every sample can be represented by its closest K neighbors. The nearest neighbor algorithm is a method to classify every record in the data set.

The core idea of the KNN algorithm is that if most of the K closest samples in the feature space belong to a certain category, then the sample also belongs to this category and has the characteristics of samples in this category. This method only determines the classification of the samples according to the category of the nearest one or several samples.

After data preprocessing, set parameter k as an odd number. Then, calculate the distance between known points and predicted points. Lastly, rank the results from small to large and take the first k points, classify the points to be predicted into the majority category.

2.2.2 Support Vector Machines(SVM)

In the research of Liu, they tried to find the best compromise point by constantly adjusting the complexity and learning ability of the SVM model. The experimental results show that SVM not only has certain effects in the field of text classification but also has good general-

	Training	Testing	Difference
Segment	8930	2760	1461
Length	60	78	18

Table 1: The difference between 2 datasetss.

```
shape: (11544, 1659)
Do you remember the movie : " The day after tomorrow " ? It reminds me of what is happening with Hurricane Sandy .
[0 2 0 ... 0 0 0]
Good Hurricane Sandy ' s photo reminds me of the independence day movie id4 sandy
[0 0 0 ... 0 0 0]
Scary shit hurricane NY
[0 0 0 ... 0 0 0]
My fave place in the world nyc hurricane sandy statueofliberty
[0 0 0 ... 0 0 0]
```

Figure 9: Selected 1*1659 vector.

ization ability [10]. Considering the difference between testing data and training data, this method is a good model to classify tweets. The difference between the 2 datasets is shown in the table 1, the segment means the number of different words in the dataset and the length means the average length of each sentence in the dataset.

2.2.2.1 Feature Selection

N-gram is an algorithm based on statistical language models. The model is based on the assumption that the occurrence of the NTH word is only related to the previous n-1 word and not to any other word, and the probability of the whole sentence is the product of the probability of occurrence of each word.

Still use CountVectorizer from sklearn to implement Bag of words model in this task. Every tweet text will transform into a 1*1659 text vector as shown in figure 9.

2.2.2.2 Dimensionality Reduction

Similarly, the data dimension after feature extraction is too high. Use ISOMAP to reduce the dimension of data.

Isomap is a non-iterative global optimization algorithm derived from the MDS algorithm. Isomap is an isometric mapping algorithm, that is, the distance between two points after dimensionality reduction remains the same, which is geodesic distance.

ISOMAP dimension reduction can be easily realized by calling the method of Sklearn library. The dimension reduction scatter diagram of five hundred random points is shown as the figure10.

2.2.2.3 Algorithm Introduction

Support Vector Machine (SVM) is a sort of generalized linear classifier that classifies data by the supervised learning method. Its decision boundary is the maximum-margin hyperplane solved for the learning sample.

As shown in the figure 11, wx+b=0 is the classification hyperplane. There is an infinite classification hyperplane for a linear separatable dataset, but the classification hyperplane with the largest geometric spacing is unique.

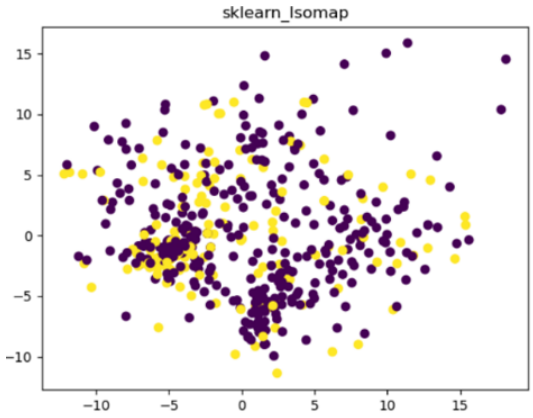

Figure 10: The dimension reduction scatter diagram of five hundred random points.

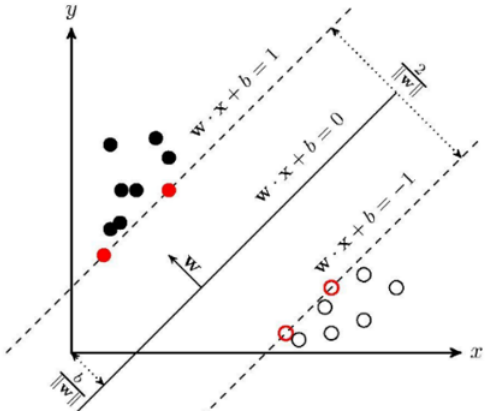

Figure 11: SVM

For the linearly indivisible samples in the finite-dimensional vector space, we map them to the vector space of higher dimensions and then learn support vector machine by the way of spacing maximization, which is nonlinear SVM.

But how to implement this? Generally kernel SVM can solve this problem,it uses kernel tricks that allow us to work in the input space rather than dealing with the potentially high-dimensional, or even theoretically infinite-dimensional, feature space [11].

2.2.3 Multinomial Naive Bayes

Yang proposed a method of categorizing the sentiment class of hotel reviews by multinomial Naive Bayes. They not only achieved good performance but also found that classification speed is fast [12], which fits nicely into tweet categories.

2.2.3.1 Feature selection

TF-IDF is a commonly used weighting technique for information retrieval and data mining.

When TF(word frequency) and IDF(reverse docu-

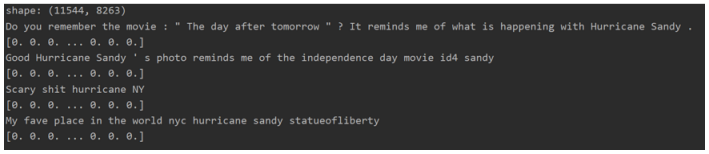

Figure 12: Selected 1*8263 vector

ment frequency) is obtained, the two words are multiplied to obtain the TF-IDF value of a word. The larger the TF-IDF of a word in the article is, the higher the importance of the word in the article will be generally. Therefore, by calculating the TF-IDF of each word in the article, the first few words are the keywords of the article.

Implement this technique with the help of sklearn in this research. After feature extraction, every text will transform into a 1*8263 text vector as shown in figure 12.

2.2.3.2 Dimensionality Reduction

This algorithm can also use Locally linear embedding since the data dimension is close to the dimension extracted from the bag of words model.

2.2.3.3 Algorithm Introduction

Multinomial Naive Bayes is one of the famous Bayes algorithms. Multinomial distribution source from polynomial experiments in statistics, which can be explained as trials involve n replicates, each with a different possible result. The probability of a particular outcome occurring in any given experiment is constant.

Bayes' theorem states that for events A and B, the probability relation between them satisfies equation (1).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

In general, the probability of event A under event B is not the same as the probability of time B under event A, but there is A definite relationship between the two, which can be described by Bayes' theorem. In the task of data classification, a new formula can show the probability relation preferably, suppose X is the feature of the data and C is the category of the data as the equation (2).

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{2}$$

Given a text whose characteristics are X, the probability that this text belongs to category C is going to calculate with the help of the training set.

2.2.4 Multilayer Perceptron

Suyash compared the results of several classical models and deep learning models in various datasets, the experimental result shows Multilayer perceptron performed

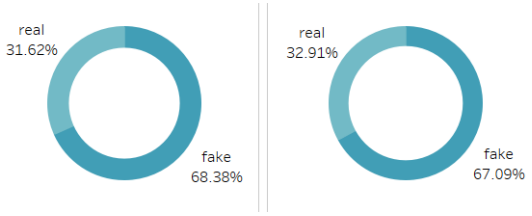

Figure 13: The labeling ratio of 2 data sets, the left one is the training set and the right one is the testing set.

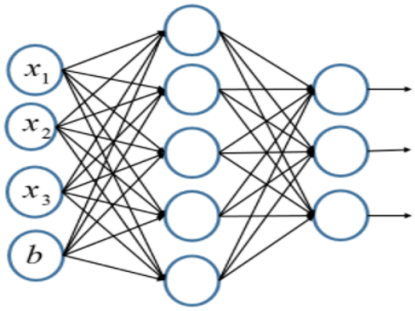

Figure 14: The structure of a multilayer perceptor

best in the dataset with uneven samples [13]. The figure13 shows the labeling ratio of the training set and testing set, which shows that the data set is not balanced.

2.2.4.1 Feature Selection

In this algorithm, still choose TF-IDF to extract features.

2.2.4.2 Dimensionality Reduction

Similarly, the dimension of data after feature extraction is too high, this algorithm would also use a Locally linear embedding model to reduce the dimension.

2.2.4.3 Algorithm Introduction

The multilayer perceptron is an upgraded version of perceptron, which combines multiple neuronal layers.

Perceptron is a linear binary classifier but can't classify non-linear data effectively. Therefore, multilayer perceptrons occurred. In theory, a multilayer network can simulate any complex function.

The structure of a multilayer perceptor is shown in the figure 14 . In the training process, a multilayer perception predicts labels by a forward pass, then updated parameters by backpropagation.

Rectified Linear Unit(RELU)

Rectified Linear Unit refers to the slope function in mathematics as shown in equation (3).

$$f(x) = \max(0, x) \tag{3}$$

Linear rectification is normally the activation function of the neuron, which defines the nonlinear output of the neuron after linear transformation.

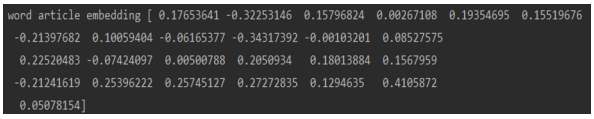

Figure 15: The word vector after training.

2.2.5 Convolutional Neural Network(CNN)

Zhang conducted several experiments on CNN with different structures under different data sets, and the results showed that CNN performed surprisingly in text classification under specific structures [14]. Considering the high-performance of CNN, choose this algorithm as the last algorithm in this research.

2.2.5.1 Feature Selection

Word2vec is an Estimator that trains word2VecModel with a series of words representing documents [15]. The model maps each word to a fixed-size vector. Word2vecmodel uses the average of each word in the document to convert the document into a vector, which can then be used as a predictive feature to calculate document similarity calculations, and so on.

Implement word2vec from genism library to train the data in this research(set vectorsize=25, which means the word vector of model training is 25 dimensions). The word vector after training is shown in the figure 15 .

2.2.5.2 Dimensionality reduction

This algorithm does not need dimensionality reduction, because the word2VEC model can extract different features according to different feature parameters.

2.2.5.3 Algorithm Introduction

Convolutional neural network(CNN) is one of the most iconic deep learning algorithms. It is widely used in the field of image classification. Compared with the traditional image CNN network, CNN for text classification has one different layer embedding layer which requires input data to be integer encoded so that each word is represented by a unique integer.

Embedding layer

The embedding layer converts positive integers (subscripts) into vectors of fixed size so that the neural network can process the text data.

Convolutional layer

The convolutional layer is composed of several convolution units, and the parameters of each convolution unit would be optimized by a backpropagation algorithm. The purpose of convolution operation is to extract various features of the input.

Pooling

Pooling, also known as subsample. The main purpose of pooling is to reduce the size of data. The most common method of pooling is max pooling and average pooling.

3 Evaluation

3.1 KNN

Strengths

1. The implementation of KNN is simple compared with other algorithms like CNN because it does not require prior training, also the core idea of KNN is easy to understand [16].
2. KNN has no assumptions for data and is not sensitive to abnormal data in the testing data compared with algorithms such as naive Bayes.
3. KNN is more suitable than other algorithms for the sample sets to be classified as more overlapping class fields. Because KNN mainly depends on the surrounding limited adjacent samples instead of the method of discriminating against the class domain to determine the category to which it belongs.

Weakness

1.KNN has a large amount of computation and memory overheads because it does not need to realize training compared to MLP with a simple structure.
2.KNN has high requirements for the quality of datasets. KNN will have the problem of misclassification if the size of the dataset is small. If the data set sample is unbalanced, the prediction accuracy of KNN for rare data points will also decrease.
3.KNN design depends on the selection of K value [17]. A bad K value will greatly reduce KNN efficiency,

3.2 SVM

Strengths

1.SVM can perform well, with high classification accuracy and strong generalization ability face to the small sample datasets [18].
2.SVM will not have the problem of local optimization in the training process compared with MLP and CNN [19]. Therefore, it's much easier to train an SVM model.
3.SVM algorithm makes the classification problem from the Bayesian algorithm from deduction to induction into efficient inference from training samples to forecast samples, greatly simplifying the classification problem.

Weakness

1.SVM scalability is not strong. Although the present experiment deals with dichotomies, it may be used to deal with multi-dichotomies such as text emotion in the future. Unfortunately, purely SVM is not ideal for handling multiple classification problems [20].
2.SVM is not suitable for big data problems compared with CNN and MLP. Because it has a large amount of computation [21] when it is dealing with a lot of sample categories, the kernel needs to be mapped to very high dimensions.
3.SVM needs better feature processing for data. SVM

could be sensitive to the absence of data features. Fewer data features may greatly reduce the effectiveness of SVM.

3.3 Multinomial Naive Bayes

Strengths

1.Multinomial Naive Bayes assumes the attributes of datasets are independent of each other, so it could achieve stable performance in different classification tasks, which allows Bayes classification to be used in many areas.
2.The core idea of Multinomial Naive Bayes is sourced from classical mathematical theory. It is simple but very robust. Therefore, Multinomial Naive Bayes is easy to implement and is not sensitive to abnormal data.
3.Multinomial Naïve Bayes doesn't need complicated computation, it can get the results quickly [22]. So, it can be applied to predict real-time applications.

Weakness

1.The condition of attribute independence is also the deficiency of naive Bayes classifiers. In many cases, it is difficult to satisfy the independence of the attributes of the data set, because there are often correlations between the attributes of the data set. Therefore, Multinomial Naïve Bayes can't perform well on many tasks in practice compared with other algorithms like MLP [23].
2.Multinomial Naïve Bayes can only be used in discrete data, it can't be used in continuous data. For those continuous datasets, Multinomial Naïve Bayes needs more complex data processing compared with other algorithms like KNN.
3.For some datasets with data scarcity, Multinomial Naïve Bayes needs to predict likelihood value by a frequentist approach, but the data scarcity could result in probabilities going towards zero or one, which makes the result unstable. Even though this situation can be solved with a smooth operation(e.g. as in sklearn) [24], it would also make Multinomial Naïve Bayes more complicated.

3.4 Multilayer perceptron

Strengths

1.Multilayer perceptron can perform well in the classification of both non-linear data and linear data. The great power of multilayer perceptron is the activation function, the combination of activation function and linear regression allowed multilayer perceptron can be applied in many areas that include regression and classification.
2.Another strength of multilayer perceptron is high-performance, it can perform better than some traditional algorithms like SVM in tasks of classification [25].
3.Multilayer perceptron has self-organization and adaptive ability, which allowed it to deal with some problems of complex environmental information, unclear knowledge background, and unclear inference rules like medical diagnosis.

Weakness

1.Multilayer perceptrons also have the disadvantage of being uninterpretable like KNN.
2.Training a good multilayer perceptron takes a lot of time. There are many problems like over-fitting or under-fitting in the training process. It may take a lot of time to mitigate these problems.
3.Multilayer perceptrons need a massive amount of data. It needs a huge amount of data to update the parameters of each layer, so it could not perform well in datasets with small sizes.

3.5 Convolutional Neural Network

Strengths

1.The main strength of CNN is its high performance. CNN with a simple structure can achieve good results on multiple data sets through slight parameter adjustment [6].
2.CNN could deal with high-dimension data efficiently compared with some algorithms like SVM due to parameter sharing and sparsity of connections (In each layer, every output value only depends on a small number of inputs.)
3.Another strength of the Convolutional Neural Network is deep. Deep means it has a deep layer structure. Multiple hidden layers enable the model to have excellent feature learning ability, and the learned data can better reflect the essential characteristics of the data, which is conducive to visualization or classification [26].

Weakness

1.Training on a good CNN is also time-consuming. In addition to the common training problems, the adjustment of network structure can also take a lot of time.
2.CNN is expensive. Building a good CNN also needs a strong hash rate. Therefore, expensive GPUs are essential.
3.Deep also could be a drawback, although deep layer could make the model perform well, it could take longer to compute compared with traditional algorithms.

3.6 Discussion

These five algorithms either fit the characteristics of the data set or have very good performance. Among them, KNN has a high requirement about the datasets but the dataset in this research has some imbalanced problems of labels even after processing so the accuracy may not be as high as expected. In addition, considering the large amount of computation of KNN, the processing time may be higher than other algorithms. Therefore, KNN is the worst algorithm of them. Then while SVM may perform well on the current data set, it is likely to perform

worse as the data increases and the problem expands. so the SVM is the fourth algorithm.

There is no doubt that CNN can achieve a good performance in this task, but the training time could be longer than other algorithms considering its complex structure. Another serious problem is expensive, in the face of exorbitant consumption, even if the performance is good, it also appears a little unsatisfactory, so CNN is the third algorithm. Similarly, MLP could achieve a good performance like CNN in this task, but the structure is simpler, so the training process could be less. Also, the adaptive ability of MLP may surprise people in practice. The only downside is that the training time and processing time may still belong, so the MLP could be the second algorithm. The best algorithm is Multinomial Naive Bayes. It has the advantage of being simple to implement and fast to process due to the principle being simple. Beyond that, even though datasets in this research could not perfectly satisfy the requirement(independence of the attributes) of Multinomial Naïve Bayes, the performance of the Multinomial Naïve Bayes maybe still good after the data set is preprocessed and the appropriate features are processed. Maybe these operations could make the model more complex, but overall it's still a simple algorithm.

To sum up, the final ranking is 1. Multinomial Naïve Bayes 2. Multilayer perceptron 3. Convolutional neural network 4. SVM 5. KNN.

4 Conclusion

In this research, five machine learning algorithms are proposed for text classification based on the data set of the Mediaeval 2015 in the face of the phenomenon of fake news flooding the Internet. Firstly, this research analyzes the format, volume, quality, and bias of the task dataset. Then, the data is preprocessed in various ways to solve the above problems. Thirdly, five machine learning algorithms including feature processing are introduced in turn, and dimension reduction is carried out for the data of the algorithm requiring feature dimension reduction. Finally, a critical analysis is made of these five algorithms, each of which contains three advantages and three disadvantages, and the final order is listed according to the comparison of the above characteristics: 1. Multinomial Naïve Bayes 2. Multilayer perceptron 3. Convolutional neural network 4. SVM 5. KNN.

Through the writing of this report, I learned the importance of data analysis. Data analysis can uncover characteristics and problems in the data set, which enable researchers to conduct better data preprocessing and algorithm selection. On the other hand, the selection of algorithms in classification tasks is also very important, the best algorithm is not the one with the best performance, but the one that best suits the characteristics of the data.

There is still a lot of work to do in the future. The first is to get more data, which would make the model fit better. Second is to conduct experiments, the most suitable model can be obtained through more parameter adjustment and data testing. Finally, the data preprocessing and feature extraction of this experiment may still be incomplete, which is also one of the problems to be solved in the future.

References

[1] M Ikonomakis, Sotiris Kotsiantis, and V Tampakas. Text classification using machine learning techniques. *WSEAS transactions on computers*, 4(8):966–974, 2005.

[2] Heui Seok Lim. Improving knn based text classification with well estimated parameters. In *International Conference on Neural Information Processing*, pages 516–523. Springer, 2004.

[3] Mohamed Goudjil, Mouloud Koudil, Mouldi Bedda, and Noureddine Ghoggali. A novel active learning method using svm for text classification. *International Journal of Automation and Computing*, 15(3):290–298, 2018.

[4] Sang-Bum Kim, Hae-Chang Rim, DongSuk Yook, and Heui-Seok Lim. Effective methods for improving naive bayes text classifiers. In *Pacific rim international conference on artificial intelligence*, pages 414–423. Springer, 2002.

[5] Fouzi Harrag and Eyas El-Qawasmah. Neural network for arabic text classification. In *2009 Second International Conference on the Applications of Digital Information and Web Technologies*, pages 778–783, 2009.

[6] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751, 2014.

[7] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.

[8] Khushbu Khamar. Short text classification using knn based on distance function. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4):1916–1919, 2013.

[9] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

[10] Zhijie Liu, Xueqiang Lv, Kun Liu, and Shuicai Shi. Study on svm compared with the other text classification methods. In *2010 Second international workshop on education technology and computer science*, volume 1, pages 219–222. IEEE, 2010.

[11] MN Murty and Rashmi Raghava. Kernel-based svm. In *Support vector machines and perceptrons*, pages 57–67. Springer, 2016.

[12] Yang Zhirui and Li Chunyan. Analysis of sentiment classification of hotel reviews based on multinomial naive bayes. In *2020 The 11th International Conference on E-business, Management and Economics*, pages 11–14, 2020.

[13] Suyash Lakhotia and Xavier Bresson. An experimental comparison of text classification techniques. In *2018 International Conference on Cyberworlds (CW)*, pages 58–65, 2018.

[14] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.

[15] Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23(1):155–162, 2017.

[16] Shichao Zhang, Xuelong Li, Ming Zong, Xiaofeng Zhu, and Debo Cheng. Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(3):1–19, 2017.

[17] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 986–996. Springer, 2003.

[18] Roman M Balabin and Ekaterina I Lomakina. Support vector machine regression (svr/ls-svm)—an alternative to neural networks (ann) for analytical chemistry? comparison of nonlinear methods on near infrared (nir) spectroscopy data. *Analyst*, 136(8):1703–1712, 2011.

[19] Mariette Awad and Rahul Khanna. Support vector machines for classification. In *Efficient Learning Machines*, pages 39–66. Springer, 2015.

[20] Eddy Mayoraz and Ethem Alpaydin. Support vector machines for multi-class classification. In *International Work-Conference on Artificial Neural Networks*, pages 833–842. Springer, 1999.

[21] Liliya Demidova, Evgeny Nikulchev, and Yulia Sokolova. Big data classification using the svm classifiers with the modified particle swarm optimization and the svm ensembles. *International Journal of Advanced Computer Science and Applications*, 7(5):294–312, 2016.

[22] Muhammad Abbas, K Ali Memon, A Aleem Jamali, Saleemullah Memon, and Anees Ahmed. Multinomial naive bayes classification model for sentiment analysis. *IJCSNS Int. J. Comput. Sci. Netw. Secur*, 19(3):62–67, 2019.

[23] Tej Bahadur Shahi and Ashok Kumar Pant. Nepali news classification using naïve bayes, support vector machines and neural networks. In *2018 International Conference on Communication information and Computing Technology (ICCICT)*, pages 1–5, 2018.

[24] Pedro Domingos and Michael Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29(2):103–130, 1997.

[25] Kwokleung Chan, Te-Won Lee, P.A. Sample, M.H. Goldbaum, R.N. Weinreb, and T.J. Sejnowski. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transactions on Biomedical Engineering*, 49(9):963–974, 2002.

[26] Xue-Wen Chen and Xiaotong Lin. Big data deep learning: Challenges and perspectives. *IEEE Access*, 2:514–525, 2014.