Fake Product Review Detection And Elimination using Opinion Mining

A.Thilagavathy, Sneha M, Shree Lakshmi R, Yuvanthika S
Department of Computer Science, R.M.K Engineering College, Tamil Nadu, India

Abstract - Identification and removal of fake reviews and its removal from the dataset provided using the supervised machine learning algorithm and natural language processing techniques(NLP) based on a vast variety of aspects. In this proposed paper, we trained the counterfeit review dataset by the process of applying two different machine learning algorithm models to identify the genuineness of the given dataset. The presence of counterfeit reviews of the products available on many E-commerce websites are mostly influencing the customers to buy those products and profit for those products is probably dependent on the reviews of those products. The products of the company were trusted before making a purchase. Hence these counterfeit reviews must be noticed so that large E-commerce companies like Flipkart, Amazon, Myntra, etc. can address this issue so that fraudsters and fraudulent critics are taken out, sustaining users' credibility in shopping sites. This approach may be utilized for websites and apps with relatively few consumers, estimating the authenticity of reviews so that online businesses can respond to them suitably. This model is developed using Naïve Bayes and TF-IDF (term frequency-inverse document frequency)Vectorizer. To detect spam reviews on a website or application instantly, one can make use of these models. However, effectively countering spammers requires a sophisticated model that has to undergo training on a large dataset of millions of reviews. In this work ”amazon Yelp dataset” a limited dataset is utilized to train the models on a small scale, but it can be expanded to achieve greater accuracy and authenticity in the reviews.

Index Terms – Opinion Mining, Data Preprocessing, Supervised Machine Learning Algorithm.

I. Introduction

The trend of people giving reviews for the product they are buying online has become a day-to-day activity nowadays. Based on the feedback consumers are buying products through various e-commerce websites. But when the reviews given by the critics are counterfeit there is no way that the consumers would not know the authenticity of the reviews provided by the critics to the customers. So consumers are being manipulated to buy a product that is not a trustworthy product. The task is straightforward but time-consuming because each review must be read and marked as a fake or ambiguous category in order to identify the true cause of the issue. By teaching a machine learning model that deals with the review section to flag a specific review as genuine or spam, this issue can be solved. The intriguing part is that this method can be used to catch spammers who didn't use the product. In order to falsely filter reviews of the product and give it a high rating, spam reviews or the use of different customer ids may be used. This can be filtered by looking at how often words like "awesome," "so good," "fantastic," etc. are used. This encourages us to create a system that uses a review's text and rating information to identify fake customer reviews of a product. By utilizing machine learning models, the honesty value and measure of a fake review will be determined. An algorithm could be used to monitor customer reviews by extracting topics and sentiments from online reviews, and it would also filter out fake reviews.

Hence the issue of fake review identification and removal requires huge data to train and be effective with added domain knowledge such as sarcasm sentences used by users to show their dissent towards the product, sometimes the product is good but not the delivery or the packing which affects the review classification. Here, an NLP technique is used to identify such reviews instead of misclassification to a negative review as in sentiment analysis. To remove unwanted or outdated product reviews include data pre-processing. Because the number of users on these websites/applications is growing daily, companies like Twitter, WhatsApp, and Facebook use sentiment analysis to detect fake news, and harmful/derogatory posts, and ban such users/organizations. This research aims to create an environment of online E-commerce where consumers can develop trust in a platform where the products they buy are genuine and feedback posted there are true and are checked regularly by the company. In addition, businesses in the e-commerce (Flipkart, Amazon), logistics, travel (Trip Advisor), job search (LinkedIn, Glassdoor), and food (Swiggy, Zomato) sectors use algorithms to combat spammers who trick customers into purchasing subpar goods and services by posting false reviews. And the users need to be alerted of the spammer like “not verified profile” hence users need not worry about such false users. Manual labeling of the reviews is practically time-consuming and less effective. So a supervised learning model is used for labeling the reviews and then predicting the label is not feasible. The Naïve Bayes and TF-IDF vectorizer methods have been utilized to identify and remove fake reviews. The fake review detection problem is addressed fairly and helps consumers to view authenticated reviews.

II. Literature Review

In [1], a preliminary study is conducted on designing a smart Learning Management System (LMS) for online learning that incorporates Natural Language Processing techniques. This study is based on a Systematic Literature Review (SLR) focused on Recommender Systems (RS). In [2], two Machine Learning (ML) models are applied to train a dataset of fake reviews in order to predict their authenticity. In [3] proposes an algorithm to track customer reviews and extract topic and sentiment information from online reviews. The algorithm is also capable of identifying and blocking fake reviews.

The proposed system in [4], called ICF++, is designed to measure a review's honesty, the reviewer's trustworthiness, and the product's reliability. In [5] examines review-centric features proposed for detecting fake reviews, with a

particular focus on approaches that employ supervised machine learning techniques. In [6] expands on a recently proposed opinion spam detection method that uses n-gram techniques by introducing feature selection and different ways of representing opinions.

In [7] suggests a novel and robust system for detecting spam reviews that effectively uses three features: (i) the sentiment of the review and its comments, (ii) content-based factors, and (iii) rating deviation. The purpose of [8] is to serve as a literature review for beginners and a survey for identifying opportunities in the field. In [9] proposes a Fake Product Review Monitoring and Removal System (FPRMS) that uses an Intelligent Interface and Uniform Resource Locators (URLs) to remove fake reviews and provide users with genuine reviews and ratings.

In [10], analyzes Yelp's filtered reviews to understand its filtering algorithm and concludes that it is reasonable and linked to unusual spamming behaviors. In [11], proposes a holistic approach called SPEAGLE that utilizes metadata such as text, timestamp, and rating, as well as relational data, to identify suspicious reviews, users, and products targeted by spam. In [12] aims to develop a machine learning model that can distinguish between genuine and fake reviews in Yelp's dataset. In [13] uses Naïve Bayes and Logistic Regression to classify Twitter reviews and assess the algorithms' performance based on accuracy, precision, and throughput.

III. METHODOLOGY

By utilizing three feature extraction techniques - Naive Bayes, TF-IDF Vectorizer, and taken dataset. We have created fake review identification. this contains the kaggle.com dataset, which includes the features listed below:

Title: the title of a news story Author: the creator of the news article.
Text: the article's text; it may be incomplete.
Label: a label indicating that the article may not be reliable
1: Untrustworthy
0: Dependable

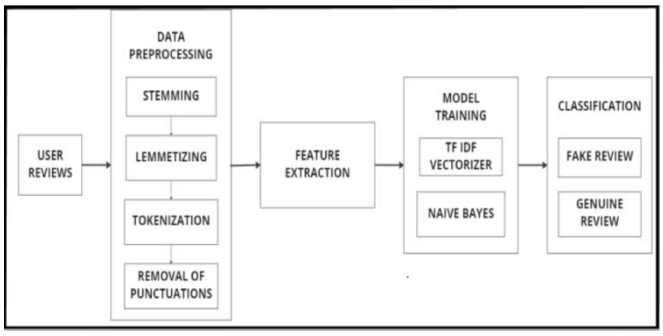

Figure 1: System Architecture

Data Pre-Processing : Processing and refining the data by removal of irrelevant and redundant information as well as noisy and unreliable data from the view dataset. The first step in the proposed approach is data preprocessing; one of the essential steps in machine learning approaches. Data preprocessing is a critical activity as the world data is never appropriate to be used. A sequence of preprocessing steps have been used in this work to prepare the raw data of the Yelp dataset for computational activities. This can be summarized as follows

Step 1: Sentence tokenization

The entire review is given as input and it is tokenized into sentences using the NLTK package. Tokenization is one of the most common natural language processing techniques. It is a basic step before applying any other preprocessing techniques. The text is divided into individual words called tokens. For example, if we have a sentence (“wearing helmets is a must for pedal cyclists”), tokenization will divide it into the following tokens (“wearing” , “helmets” , “is” , “a”, “must”, “for” , “pedal” , “cyclists”)

Step 2: Removal of punctuation marks

Punctuation marks used at the starting and ending of the reviews are removed along with additional whitespaces. Stop words are the words which are used the most yet they hold no value. Common examples of the stop words are (an, a, the, this). In this paper, all data are cleaned from stop words before going forward in the fake reviews detection process.

Step 3: Word Tokenization

Each individual review is tokenized into words and stored in a list for easier retrieval.

Step 4: Removal of stop words

Affixes are removed from the stem. For example, the stem of "cooking" is "cook", and the stemming algorithm knows that the "ing" suffix can be removed.

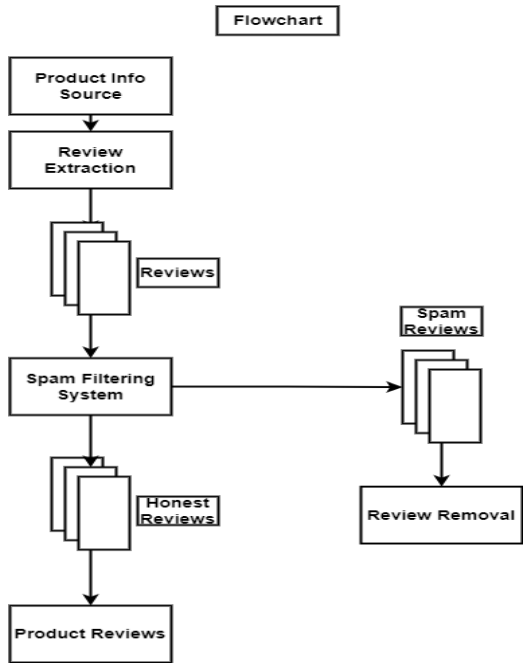

Figure 2: Flowchart for fake review detection

We initialize a Naive Bayes Classifier to fit the model after using TF-IDF Vectorizer to translate our text strings into numerical representations. In conclusion, our model's performance is indicated by the accuracy score and confusion

matrix. TF-ID. A popular algorithm for converting text into intelligible numerical representations is called vectorizer. Based on occurrence, it is used to extract features from text strings. We presume that a word's relevance in the provided text will increase with its frequency of repetition. We normalize a word's frequency based on the volume of the document and refer to this as term frequency.

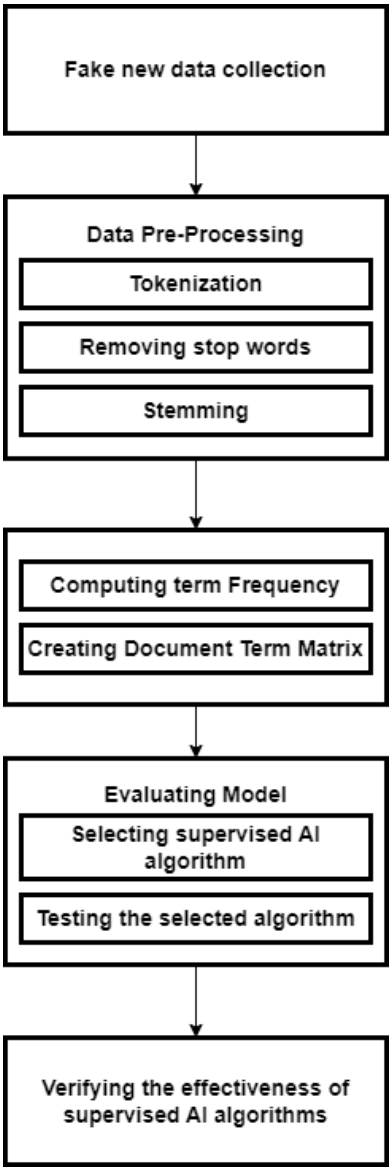

Figure 3: Workflow for identifying and analyzing reviews using different algorithms

Calculated definition:

$$TF(w) = \frac{doc.count}{total\_words\_in\_the\_document}$$

Each term is weighted equally when determining term frequency. Words that are often used in papers could be less helpful in determining the meaning of the document because of their high frequency. The weights of more important words may be suppressed by such words as "a," "the," and similar ones. TF is discounted by the inverse document factor in order to lessen this impact.

$$IDF(w) = log\left(\frac{total\_number\_of\_documents}{number\_of\_documents\_containing\_word}\right)$$

Then, by multiplying TF and IDF, one may obtain TF-IDF. The TF-IDF score would increase for terms that are more significant.

$$TF - IDF(w) = TF(w) * IDF(w)$$

Naive Bayes classifiers are a subset of simple machine learning in artificial intelligence. Using multinomial NB and pipelining ideas, the widely used Naive Bayes algorithm determines whether the news is accurate and true. There are several strategies for training these classifiers that concentrate on common principles, therefore this is not the only algorithm available.

Naive Bayes can be used to determine whether the news is phony or authentic. It is a form of an algorithm that is used to categorize texts. The correctness of the news is determined using the Bayes theorem after the use of tokens is associated with the news that may or may not be false. The recipe for being naive is as follows: The likelihood of the prior occurrence is used in Bayes classification, which contrasts it with the current event. A final calculation is made to determine the overall likelihood of the news when compared to the dataset after calculating each and every probability of the occurrence. As a result, by computing the overall likelihood, we may determine an approximation of the value and determine if the news is true or false.

$$P(B) = P(A) - \frac{P(A)}{P(B)} \qquad (1)$$

Finding the probability of an event, A when event B is true.

$$P(A) = PRIOR\ PROBABILITY$$

$$P(A|B) = POSTERIOR\ PROBABILITY$$

Finding probability:

$$P(B1) = P(A1||B1).P(A2||B1).P(A3||B1) \qquad (2)$$

$$P(B2) = P(A1||B2).P(A2||B2).P(A3||B2) \qquad (3)$$

If the probability is 0

$$P(Word) = \frac{Word\ count\ +\ 1}{Total\ no.of\ words\ +\ No.of\ unique\ words}$$

Consequently, one can determine the news' accuracy by applying this method.

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two

different categories that are classified using a decision boundary or hyperplane:

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

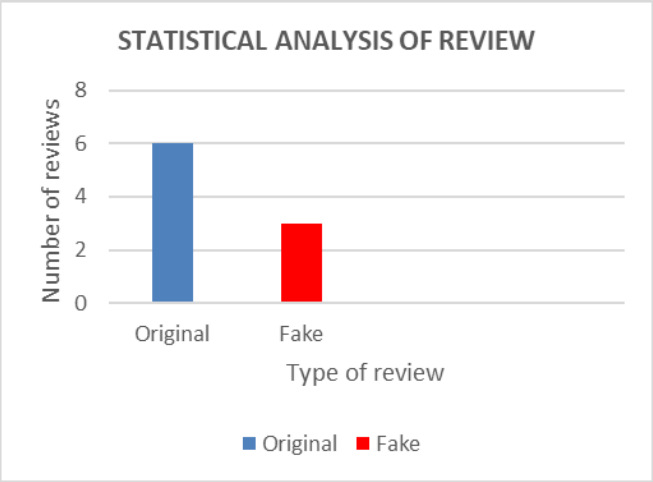

Figure 4: Statistical analysis of review

IV. Results & Discussion

This model has examined the detection of false reviews, which are increasingly common on websites and social media platforms. Our model was trained using text processing and Naive Bayes. So, by utilizing tools for machine learning we can draw the conclusion that any news from a large or small dataset can be categorized as being fake or not fake with the help of prior data set values in a shorter amount of time, enabling the user to trust in specific reviews that emerge on social media or from other sources.

TABLE 1: Summary of the Dataset	
Total number of reviews	5853 reviews
Number of fake reviews	1144 reviews
Number of real reviews	4709 reviews
Number of distinct reviews	102739 words
Total number of tokens	103052 tokens
The maximum review length	875 words
The minimum review length	4 words
The average review length	439.5 word

Users and the social environment are severely harmed by the dissemination of false information. Because the fake review is intended to mislead the user, it is challenging to identify them in the first place. False information is disseminated through a variety of channels, which disrupts society and the lives of its citizens. Finding the source of the false information and putting an end to its dissemination on social media and online platforms would be future improvements. In order to stop those who are attempting to mislead the public, it would also be able to track down and identify the sources of false information. Also, they would locate the social media accounts of those propagating rumors and fake news so they could stop them before it's too late. These are all things that can infuse society with goodness and a healthy way of living.The formula for calculating the conditional probability of the fact, that review is fake given that it contains some specific word looks as following:

$$\Pr(F|W) = \Pr(W|F) \cdot \frac{\Pr(F)}{\Pr(W|F) \cdot \Pr(F) + \Pr(W|T) \cdot \Pr(T)}$$

We then take the logarithm (with base 2) of the inverse frequency of the paper. So the if of the term t becomes:

$$idf(t) = log(N/df(t))$$

The TF-IDF score would increase for terms that are more significant.

$$TF - IDF(w) = TF(w) * IDF(w)$$

Hence, The proposed system provides the accuracy around 83-89% in various test case scenarios.

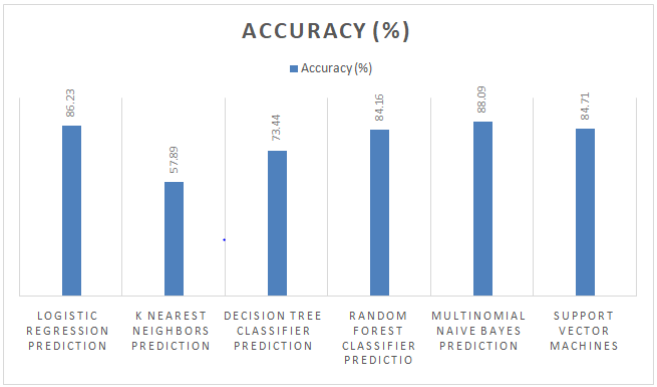

Figure 5 : Accuracy percentage of different algorithms

Users and the social environment are severely harmed by the dissemination of false information. Because the fake review is intended to mislead the user, it is challenging to identify them in the first place. False information is disseminated through a variety of channels, which disrupts society and the lives of its citizen.

$$Precision = \frac{TP}{TP+FP} \times 10$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100$$

$$Specificity = \frac{TN}{TN+FP} \times 100$$

$$F1 - Score = 2 \times \frac{precision \times sensitivity}{precision + sensitivity} \times 100$$

where true negative (TN) represents the total number of samples that were effectively predicted as truthful reviews by the classifier. False negative (FN) represents the total number of samples that were incorrectly classified as fake reviews. True Positive (TP) denotes the total number of samples that were successfully classified as fake reviews. False Positive (FP) is the sum of samples that were incorrectly categorized as truthful reviews.

Figure 6: Confusion matrix for SVM (left) and NB (right)

By comparing the classification results of the proposed classifiers, the RF classifier provided the best performance at detecting fake reviews and outperformed other classifiers with a 86% accuracy and F1- score metric. The sample classifications through RF were based on the majority voting of multidecision trees. The NB classifier provided equal numbers of positive and negative samples and had better results than SVM and Decision Tree classifiers with a 88% sensitivity metric. The naïve Bayes classifier had the highest misclassification, yielding an 88% accuracy and F1-score metric.

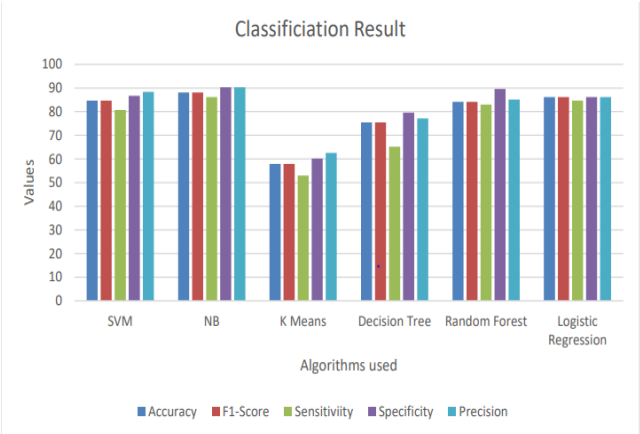

Figure 7: Visualization of Classification Result

The below table provides a summary of the dataset by segregating the total number of reviews ,Number of fake reviews , Number of real reviews Number of distinct reviews ,total number of reviews , maximum , review length ,minimum review length and the average review length. With this table it is easy to understand the dataset used

V. Conclusion

This paper demonstrates how Artificial Intelligence and Internet of Things technology, and most prevalent machine learning techniques, can be deployed to tackle issues that have plagued the blind, deaf, and deaf individuals for years. Giving the deaf a voice and assisting them in understanding what those around them are saying may be possible with automatic speech recognition. The blind could be guided while walking with the help of live image transmission and onboard object recognition, effectively giving them an eye to see what is going on around them. Smart glasses would assist in resolving the issues of many blind, deaf, and dumb people who lack confidence in speaking and even assisting them in various activities. The blind and deaf might both benefit from this efficient, low-cost technology's ability to see and hear.

VI . References

[1] D. F. Murad, Y. Heryadi, B. D. Wijanarko, S. M. Isa and W. Budiharto, "Recommendation System for Smart LMS Using Machine Learning: A Literature Review," 2018 International Conference on Computing, Engineering, and Design (ICCED), Bangkok, Thailand, 2018, pp. 113-118, doi: 10.1109/ICCED.2018.00031.

[2] S. M. Anas and S. Kumari, "Opinion Mining based Fake Product review Monitoring and Removal System," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 985-988, doi: 10.1109/ICICT50816.2021.9358716.

[3] Jain, Piyush & Chheda, Karan & Lade, Mihir. (2019). Fake Product Review Monitoring System. International Journal of Trend in Scientific Research and Development. Volume-3. 105-107. 10.31142/ijtsrd21644.

[4] Wahyuni, Eka & Djunaidy, Arif. (2016). Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework. MATEC Web of Conferences. 58. 03003. 10.1051/matecconf/20165803003.

[5] Kashid, Aishwarya & Lalwani, Ankita & Gaikawad, Saniksha & Patil, Rajal & Sonkamble, Rahul & More, Shivaprasad. (2021). Fake Review Detection System Using Machine Learning.

[6] R. Patel and P. Thakkar, "Opinion Spam Detection Using Feature Selection," 2014 International Conference on Computational Intelligence and Communication Networks, Bhopal, India, 2014, pp. 560-564, doi: 10.1109/CICN.2014.127.

[7] Saumya, S., Singh, J.P. Detection of spam reviews: a sentiment analysis approach. CSIT 6, 137–148 (2018). https://doi.org/10.1007/s40012-018-0193-0

[8] N. Sodera and A. Kumar, "Open problems in recommender systems diversity," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 82-87, doi: 10.1109/CCAA.2017.8229776.

[9] Ata-Ur-Rehman et al., "Intelligent Interface for Fake Product Review Monitoring and Removal," 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 2019, pp. 1-6, doi: 10.1109/ICEEE.2019.8884529.

[10] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2021). What Yelp Fake Review Filter Might Be Doing?. Proceedings of the International AAAI Conference on Web and Social Media, 7(1), 409-418.

[11] 2015. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA.

[12] A. Sihombing and A. C. M. Fong, "Fake Review Detection on Yelp Dataset Using Classification Techniques in Machine Learning," 2019 International Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 64-68, doi: 10.1109/IC3I46837.2019.9055644

[13] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," 2017 International Conference on

Computer Communication and Informatics (ICCCI), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICCCI.2017.8117734.