Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Sentiment analysis of social media discourse on public perception of online courier services in Saudi Arabia using machine learning

Mohamed Shenify[a*]

[a] *College of Computing and Information, Al-Baha University, Saudi Arabia*

C H R O N I C L E

Article history:
Received: July 10, 2024
Received in revised format: August 1, 2024
Accepted: August 5, 2024
Available online: August 10, 2024

Keywords:
Social media
Sentiment analysis
Machine learning
Decision tree
SVM
Online courier services

A B S T R A C T

The Kingdom of Saudi Arabia has witnessed a significant surge in online shopping in recent years, fueled by factors like growing internet penetration, smartphone adoption, and government initiatives supporting e-commerce growth. This rise in online activity has led to a corresponding increase in the utilization of online courier services, playing a crucial role in ensuring timely and efficient delivery of goods In this context, understanding public perception of online courier services becomes crucial for businesses to improve their offerings, address customer concerns, and maintain a competitive edge. Social media platforms have emerged as a valuable source of customer feedback and user-generated content, offering insights into customer experiences and opinions. This paper presents a sentiment analysis on online couriers in Saudi Arabia using natural language processing techniques combined with Decision Tree and Support Vector Machine (SVM) classifiers of machine learning. A dataset on customers' sentiments was created by a crawling process from X social media. Both classifiers perform well, with Decision Tree classifier performs slightly better on accuracy, i.e. 95.01% compared to 93.60% of the Support Vector Machine. Other metrics support the robustness of the classification.

1. Introduction

The Kingdom of Saudi Arabia has witnessed a significant surge in online shopping in recent years, fueled by factors like growing internet penetration, smartphone adoption, and government initiatives supporting e-commerce growth. This rise in online activity has led to a corresponding increase in the utilization of online courier services, playing a crucial role in ensuring timely and efficient delivery of goods (Salem et al., 2020; Mordor, 2024). In this context, understanding public perception of online courier services becomes crucial for businesses to improve their offerings, address customer concerns, and maintain a competitive edge. Social media platforms have emerged as a valuable source of customer feedback and user-generated content, offering insights into customer experiences and opinions (Zhang et al., 2022). From the perspective of business, Saudi Arabia has specific landscape where online courier service market is experiencing rapid growth, with positive competition, which leads to the improvement of service quality and competitive pricing, the enhancement of efficiency on delivery capabilities, and the adoption of Artificial Intelligence (AI) and automation for optimizing operations (Sobaih et al., 2023). The customer feedback landscape is shifting from the traditional way where customer feedback was gathered through surveys or direct communication with businesses to the rise of social media where customers increasingly voice their opinions and experiences on media platforms. The social media

* Corresponding author.
E-mail address maalshenify@bu.edu.sa (M. Shenifya)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2024.8.002

facilitates real time feedback that allows courier providers to address any issues or concerns without any significant due. The positive and negative experiences are shared in the social media and can significantly impact the image and reputation of the courier providers as well as the customer acquisitions (Sghaier et al., 2016). Therefore, understanding public perception of online courier services in Saudi Arabia through sentiment analysis of social media discourse is crucial for businesses to gain valuable insights into customer satisfaction and identify areas for improvement, by instantaneous response to customer's concerns and foster positive brand sentiment. While sentiment analysis has become a powerful tool for analyzing public opinion, several key gaps exist in understanding public perception through this approach.

The field of sentiment analysis using AI is constantly evolving, with significant advancements in both techniques and their application to social media (Adak et al., 2022; Alsari et al., 2022: Semary et al., 2022: and Elhag et al., 2021). The state of the art of techniques in AI-powered sentiment analysis of social media include: Deep Learning Models, particularly recurrent neural networks (RNNs) and transformers, which excel at capturing complex relationships within text data; Long Short-Term Memory (LSTM) networks, which effective for analyzing sequential data like social media posts, as they can learn long-term dependencies within sentences; Transfer Learning where pretrained models like Bidirectional Encoder Representations from Transformers(BERT) are fine-tuned for specific tasks like sentiment analysis in social media, leveraging pre-existing knowledge to improve performance.

The main objective of this research is to explore public perception of online courier services in Saudi Arabia using AI-powered sentiment analysis of social media discourse. More specifically, to gain insights into the factors influencing public perception, potentially including delivery speed, service quality, customer support, and overall experience.

2. Theoretical Background and Related Works

2.1 Sentiment Analysis

Sentiment analysis is a way to understand people's opinions about something by looking for clues in their words. It can figure out if people are happy, sad, or neutral based on what they write. This can be done by examining entire documents, individual sentences, or even specific features or aspects being discussed. There are three main methods for sentiment analysis: using lists of emotional words (lexicon-based techniques), training computers to recognize patterns (machine-learning-based techniques), or a hybrid approach (Dang et al., 2020).

The very first methods used in sentiment analysis relied on dictionaries of words. These fell into two categories: predefined lists of positive and negative words, or analyzing a large group of documents to find statistical patterns. The first category uses dictionaries like SentiWordNet or WordNet. The second category uses corpus-based analysis, which looks for patterns in big collections of documents using techniques like k-nearest neighbors, conditional random fields, and hidden Markov models, thus, does not need a set list of words

Sentiment analysis primarily relies on two techniques: traditional machine learning and deep learning. Traditional methods, such as Naïve Bayes and SVMs, analyze text features like word choice and grammar to identify sentiment. However, their effectiveness is limited by the factors they consider. Deep learning models, including deep neural networks (DNNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs), often outperform traditional methods by examining entire texts or specific elements to accurately determine sentiment.

Before training a sentiment analysis model, text data must undergo cleaning, regardless of whether using deep learning or traditional machine learning. Twitter data, for instance, contains unnecessary elements such as extra spaces, punctuation, and irrelevant words. These elements are removed using tools like BeautifulSoup to focus on sentiment-bearing content. After cleaning, text is broken down into individual words, which are then simplified to their root form (lemmatization). To make these words understandable to the machine, they are converted into numerical representations using techniques such as word embedding or frequency-inverse document frequency (TF-IDF).

Word embedding is a method that converts words into numerical representations, or vectors, where similar words have similar vectors (Vaswani et al., 2017). These vectors are learned using neural networks. Popular tools like Word2Vec and GloVe create these embeddings by analyzing word occurrences within a text. Two primary techniques in Word2Vec are skip-gram and CBOW, which predict surrounding words or the center word based on context, respectively. Another technique, TF-IDF, assesses a word's significance within a document compared to others. It calculates a score based on word frequency within a document and across a collection. Words appearing frequently in a specific document but rarely in others have higher scores. The scikit-learn library provides a tool for this calculation. In natural language processing, sentiment analysis transforms text into numerical data. Deep learning models utilize features derived from text, such as word embeddings and TF-IDF, to accomplish the task.

2.2 Decision Tree

A decision tree is a machine learning model used to categorize data. It functions like a flowchart, where each decision point (node) asks a question about the data, and the possible answers determine the next step. This process continues until a final category (leaf) is reached. The leaves of the tree represent the final classification of the data point.

A decision tree algorithm constructs a tree-like model using data with features and corresponding labels. It begins by selecting the feature that best separates the data into groups based on the target label. This becomes the root of the tree. The process repeats for each subgroup, creating branches until the data is pure (all data points belong to the same category) or a specified tree depth is reached. The final tree can then classify new data by following the decision paths based on the data's features, ending at a prediction. Decision tree is powerful due to its interpretable characteristic; however, it is prone to overfitting and sensitive to feature scaling (Gulo et al., 2022). Fig. 1 illustrates a general decision tree scheme.

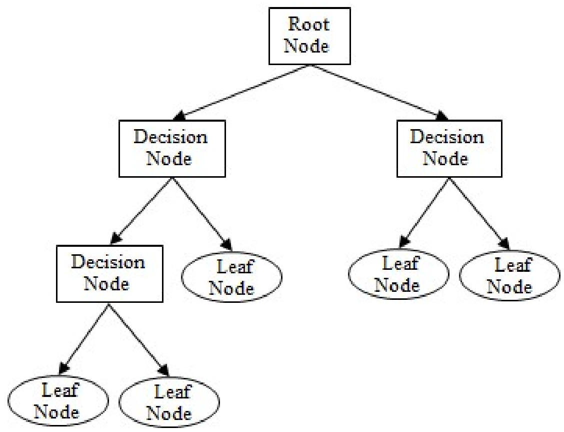

Fig. 1. General decision tree scheme

2.3 Support Vector Machine (SVM)

SVM is a machine learning technique for categorizing data. They work by finding the best dividing line (hyperplane) between different data groups. This line is positioned to maximize the distance between the closest data points from each group. This approach makes SVM robust to outliers and effective for complex datasets. SVM focuses on key data points to create the most accurate dividing line by finding the most important data points, such as cornerstones, to build the best possible hyperplane (Wang et al., 2023).

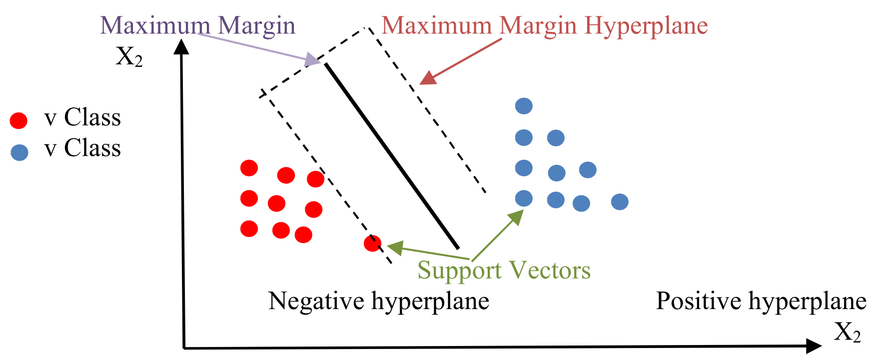

Fig. 2. Support Vector Machine (Mohammadi et al., 2021)

Fig. 2 illustrates how SVM can clearly separate data into different categories (Mohammadi et al., 2021). The core idea is to find the best possible dividing line (hyperplane) between these categories, maximizing the distance between the closest data points on each side. These critical data points are called support vectors, and they determine the hyperplane's position. SVM excels at creating this optimal separation by focusing on these support vectors. The effectiveness of an SVM often depends on selecting the appropriate kernel function. Thus, choosing the right kernel function is crucial for effective SVM applications.

2.4 Related Works

Hakami's study (Hakami, 2023) used machine learning to understand what Saudi Arabian consumers think about popular shopping apps. The research identified key factors influencing customer satisfaction and sales. Positive reviews focused on fast and reliable delivery, easy shopping experiences, product quality, competitive prices, good customer service, and suitable product sizes. On the other hand, negative feedback centered around poor service, refund issues, problems with returns, incorrect product sizes, app updates, and delivery delays. These insights can help e-commerce businesses improve their services and boost customer satisfaction. A recent study by Alsari et al. (2022) looked at how people felt about their very first Saudi cruise experiences. They gathered information from social media posts on Instagram, Snapchat, and Twitter (three platforms in total). By analyzing the emotions expressed in these posts, they aimed to understand passenger and viewer opinions. The researchers used machine learning to categorize over 1200 cleaned entries into positive or negative sentiment. Five different machine learning algorithms, i.e.: Multilayer Perceptron, Naive Bayes, Random Forest, SVM, and Voting are deployed. Interestingly, the Random Forest algorithm achieved a perfect score (100% accuracy) when analyzing oversampled data from Snapchat (data where the number of positive and negative entries was balanced). Overall, all the algorithms performed well across the different social media platforms. Most importantly, the analysis revealed that roughly 80% of the sentiment towards these cruises was positive, with only 20% being negative.

Wasiq et al., (2022) investigated what drives people in Saudi Arabia to use mobile commerce (M-commerce) services, especially during the COVID-19 pandemic. Their study looked at four main factors: individual characteristics, economic conditions, how easy it is to use M-commerce, and safety concerns due to COVID-19. They surveyed 340 M-commerce users in Saudi Arabia to gather data. Using statistical methods, they found that all four factors significantly influence customers' decisions to adopt and use M-commerce. In other words, these factors play a role in why people choose to shop using mobile commerce platforms. The study also revealed an increase in M-commerce use during the pandemic, likely due to health and safety measures requiring social distancing. A limitation of the research is that it focused on a specific set of factors, and there might be others to consider.

While, a study by Zygiaris et al., (2022) investigated how well car service companies meet customer needs after the pandemic. The researchers found that social media platforms were particularly useful for repair shops to keep customers informed and address their concerns promptly. The study used a framework called SERVQUAL to measure the connection between the quality of service and customer satisfaction. The findings showed that all aspects of service quality – empathy, reliability, assurance, responsiveness, and even the physical facilities – had a positive impact on customer satisfaction.

3. Materials and Methods

Figu. 3 shows the stages of the proposed method. The first stage, data creation starts with the data crawling from X social media platform then followed by data preprocessing to produce a clean dataset. The next step is to conduct sentiment analysis experiments using the prepared dataset. Detailed explanations of this process will be provided in the following sections.

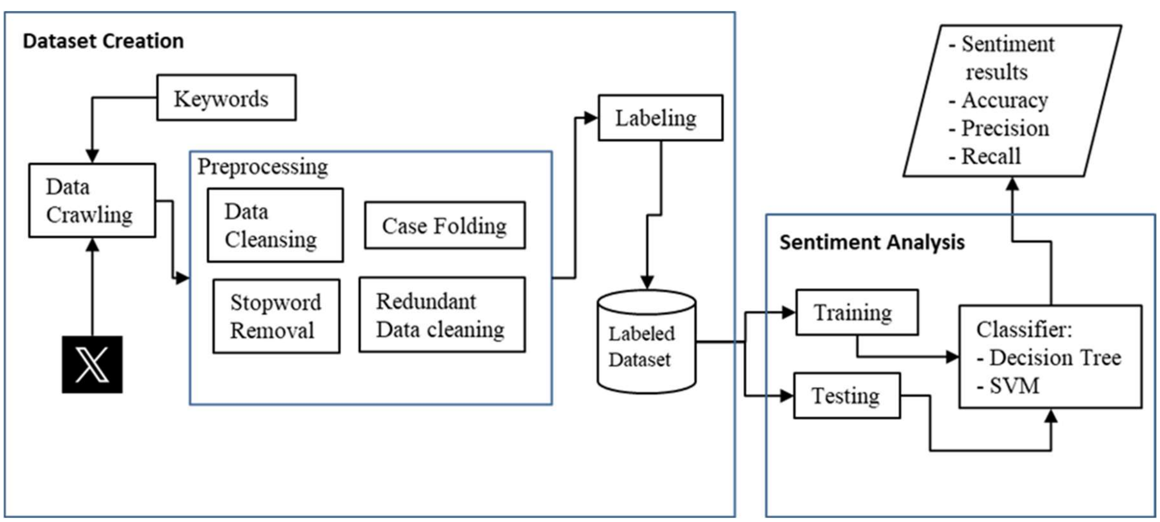

Fig. 3. Stages in the proposed method

3.1 Dataset Creation

The dataset used in this research is obtained from X (formerly known as Twitter). To implement the data crawling process, the Tweepy library of Python programming language is used to access the X (Twitter) API. Firstly, a list of X users who were targets for data collection had been determined and entered into the program. The data was taken from 100 X accounts owned by influential figures in Saudi Arabia, in order to get as much information as possible from the users who send tweets most often on Twitter. Each Twitter user took 400 tweets, resulting in a dataset with a total of 45,243 tweets from the 100 users.

3.1.1 Data Preprocessing

Preprocessing is one of the very important steps to be carried out in the sentiment analysis process because this process helps in building efficient, stable and strong data models (Prabhakar et al., 2019). Data preprocessing generally has stages consisting of data cleansing, case folding, stopwords removal, and redundant data cleaning, with the following details.

Data Cleansing: At this stage, unnecessary data is deleted, such as punctuation, hashtags, emoticons and links. Cleaning is carried out with the aim of getting better results from the labeling process because it removes all unnecessary characters such as hashtags, commas, question marks and so on which can reduce labeling performance.

Case Folding: The data obtained from X is raw data and contains non-uniform letters. To be able to carry out the next preprocessing stages, first, change the letters to lower case, so that all the letters in each tweet are uniform and can be processed and read by machines better.

Stopwords Removal: Data that has been previously cleaned at the cleansing stage still contains many stop words that are not related to the analysis process that will be carried out, so these words must be removed from the tweets. Usually these words are question words and time telling words. Stopword removal for Arabic is carried out using the literary library, then matching the words in the dataset and those in the library will be carried out At this stage, several additional words are used to accommodate the variety of new words in the dataset. These additional words will later be entered into a data dictionary in the form of an array or set, and then the dataset will match the words in the dataset with the words in the data dictionary.

Redundant data cleaning: When pulling data from X, some data contains redundant data from the same user. For example, the username "@a" sends the tweet "I like them" more than once, will create redundant records in the *dataframe*, therefore repeating tweet data from the same user will be deleted later. This deletion of redundant data is done using the 'drop_duplicates()' line of code. This method is done after importing the data that has been combined into the Pandas dataframe, then the syntax will be used in the existing Pandas *dataframe*.

3.2 Sentiment Analysis Process

3.2.1 Word Weighting

After preprocessing and labeling the data, the next step is to weight the words using TF-IDF, where the number of times a word appears in the corpus is counted. By using TF-IDF, words are not only counted how many times they appear in a corpus but how much weight the word has in the corpus based on labeling.

3.2.2 Data Splitting

After weighing the words, the dataset is split into two parts, with a specific ratio for testing and for training processes. This splitting is useful for finding the best ratio where the classifiers achieve maximum performance. Figure 4 illustrates the implementation of data splitting in the experiment.

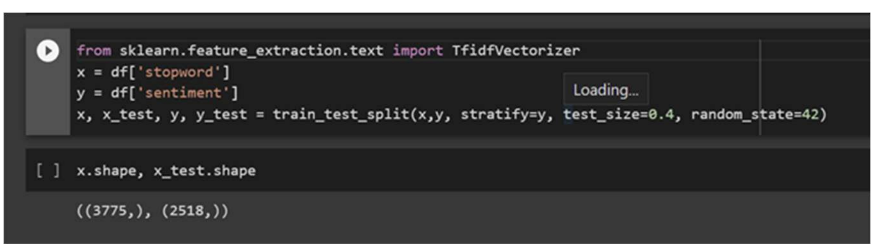

Fig. 4. Implementation of data splitting

The model building is also carried out using random states. The use of random states aims to ensure that the results in the model can be reproduced and have full control over the data.

3.2.3 Building the Model

The decision tree is constructed through a repetitive process outlined in Algorithm 1. This process involves building the tree step-by-step (Nisbet et al., 2018).

Algorithm 1. Decision Tree (Nisbet et al., 2018)

```
1. Assign all training instances to the root of the tree. Set current node to root
   node.
2. For each attribute
a. Partition all data instances at the node by the value of the attribute.
b. Compute the information gain ratio from the partitioning.
3. Identify feature that results in the greatest information gain ratio. Set this
   feature to be the splitting criterion at the current node.
   a. If the best information gain ratio is 0, tag the current node as a leaf and
      return.
4. Partition all instances according to attribute value of the best feature.
5. Denote each partition as a child node of the current node.
6. For each child node:
a. If the child node is "pure" tag it as a leaf and return.
b. If not set the child node as the current node and recourse to step 2.
```

The Sigmoid SVM (S-SVM) is adopted in this paper. The Sigmoid kernel function is represented in Eq. (1).

$$K(x, x_i) = tanh(\alpha x_i \bullet x_j + \beta) \tag{1}$$

The classifier is implemented as Algorithm 2 (Wang et al., 2023).

Algorithm 2. SVM (Wang et al., 2023)

```
1.   Input: Cleaned_Data_Set
2.   Processing:
3.   Kernal_Type = SetKernelType ()
4.   SVM_Classifier = GetSVMClassifier (Kernal_Type)
5.   Classified_Data_Set = SVM_Classifier.Train (Cleaned_Data_Set)
6.   Output: Classified_Data_Set
```

3.3 Performance Metrics

A confusion matrix is a tool used to assess the performance of a classification model. It compares the model's predictions to the actual outcomes, categorizing results into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). By analyzing these values, the model's accuracy can be determined. Additionally, several metrics, such as accuracy, precision, recall, and F1-score, can be calculated from the confusion matrix to provide a more comprehensive evaluation of the model's performance. Tuan et al., (2020) also mention the four metrics used to evaluate the performance of a system based on the confusion matrix as shown in Eq. (2) – Eq. (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\_Score = \frac{2TP}{2TP + FP + FN} \tag{5}$$

4. Results and Discussion

The models were developed using Python 3.10.1 on a high-performance personal computer (PC) equipped with a six-core Intel i7 processor, 16 gigabytes of RAM, and a 1 terabyte hard drive. The operating system is Windows 11 Home (64-bit).

4.1 Created Dataset

The limitations of the Application Programming Interface (API) provided by X contribute to very significant impact on the data crawling, due to no longer possible to retrieve data from 7 days before the day of data collection. The data was taken during 2 to 9 of March 2024. The number of tweets selected and labeled for the experiment was 6,482 records. The most appeared words are displayed in Table 1.

Table 1
The most appeared words

No.	Word in Arabic	Meaning	%	No.	Word in Arabic	Meaning	%
1	رزمه	Packet	8	19	مكتب	Office	2
2	تقدير	Estimation	8	20	اتصال	Contact	1.2
3	يساعد	Assist	6	21	حالة	Status	1.2
4	يحتاج	Need	6	22	يعبر	Express	1
5	عمل	Business	6	23	سريع	Fast	1
6	غداً	Tomorrow	5	24	بطيء	Slow	1
7	يرسل	Send	5	25	يمسك	Hold	1
8	يسلم	Deliver	5	26	يتغير	Change	0.8
9	يستلم	Receive	5	27	خدمات	Services	0.8
10	شكرًا لك	Thank you	4	28	إجابة	Response	0.8
11	الخسارة / المفقودين	Loss/missing	4	29	يكتب	Write	0.8
12	مستودع	Warehouse	4	30	يتحرك	Move	0.8
13	بورسيل	Purcell	4	31	سيء	Bad	0.6
14	متابعة	Follow up	4	32	ممتاز	Excellent	0.6
15	خطأ	Wrong	3	33	أمس	Yesterday	0.6
16	الرياض	Riyadh	3	34	أسبوع	Week	0.6
17	ساعي	Courier	2	35	منزعج	Upset	0.6
18	منزل	House	2	36	جيد	Good	0.6

The labeling process produces 4,225 positive sentiment tweets, 1,950 negative sentiment tweets, and 1 undefined tweet.

4.2 Classification Results

The confusion matrices obtained from the experimental results are presented in Table 2 and Table 3.

Table 2
The confusion matrix values from the training experiments

	90:10		80:20		70:30		60:40	
	DT	SVM	DT	SVM	DT	SVM	DT	SVM
TP	5202	5092	4547	4526	3834	3754	3470	3463
FP	631	741	638	659	703	783	419	426
TN	563	647	560	576	594	648	374	380
FN	68	94	78	83	109	135	45	46

Table 3
The confusion matrix values from the testing experiments

	90:10		80:20		70:30		60:40	
	DT	SVM	DT	SVM	DT	SVM	DT	SVM
TP	5275	573	4927	608	4210	592	3328	542
FP	558	76	258	41	327	57	561	107
TN	498	67	227	36	277	48	501	96
FN	60	9	31	5	50	9	60	11

Table 4 summarizes the four metrics for the performance of the classification, i.e.: F-Score, recall, precision, and accuracy. The 80%:20% data split for training and testing provides the best results. In this scenario, the Decision Tree classifier achieved the highest accuracy of 95.01%, while the SVM Classifier reached 93.60%.

Table 4
Classification performance result

Scenario	Accuracy		Precision		Recall		F1-Score	
	DT	SVM	DT	SVM	DT	SVM	DT	SVM
90%:10%								
Training	89.17%	87.28%	88.14%	87.76%	88%	87%	88.65%	87.71%
Testing	90.43%	88.28%	88.72%	88.61%	88%	88%	90.33%	89.93%
80%:20%								
Training	87.69%	87.28%	86.44%	86.28%	87%	87%	87.99%	87.13%
Testing	**95.01%**	**93.60%**	92.98%	93.60%	93 %	93%	94.04%	93.55%
70%:30%								
Training	84.49%	82.73%	86.05%	86.84%	87%	87%	87.46%	87.02%
Testing	95.78%	91.10%	90.80%	90.75%	91%	90%	90.77%	90.96%
60%:40%								
Training	89.22%	89.04%	88.97%	88.65%	90%	89%	89.22%	89.03%
Testing	85.56%	83.37%	84.55%	83.37%	88%	83%	89.51%	89.22%

5. Discussion

Both classifiers, the Decision Tree and SVM demonstrated high accuracy. This performance is supported by other evaluation metrics, i.e.: precision, recall, and F1-score. Additionally, there is no evidence of overfitting as the model's performance on testing data is slightly better than its training performance.

From Table 4, it can be observed that both classifiers give the best accuracy for the same scenario of data split, i.e.: 80%:20%. Decision Tree achieves significantly better performance compared to SVM, due to the fact that Decision Tree for sentiment analysis has strength in interpretability (Gulo et al., 2022). It is also relatively fast to train, especially on smaller dataset and less sensitive to outliers in the dataset. In addition, it can handle well categorical data.

n the other hand, SVMs also achieve high accuracy in sentiment analysis tasks, as the dataset is well-structured and balanced. Their ability to find the optimal hyperplane for separating classes leads to effective classification. However, the SVM model is not easily interpretable, which is difficult to understand the reasoning behind the prediction results. Selecting the optimal hyperparameters for SVM also can be a complex and time-consuming process, which impacts the classification performance, as the tuning process is crucial. Furthermore, SVM often requires more extensive data preprocessing compared to decision trees. Features might need scaling or normalization to ensure they contribute equally during training (Wang et al., 2023).

The size of the dataset used and the splitting methods might be considered limitations of this study. Thus, a larger dataset as well as better splitting algorithms should be considered for future works.

6. Conclusions

The use of Decision Tree classifier algorithm with an 80%:20% data splitting provides the highest results compared to other data splitting. This optimal condition produces the highest accuracy, of 95.01% and 93.60% for Decision Tree and SVM algorithms, respectively with good support of other metrics (precision, recall and F1-Score). Reviews of opinions posted on social media X conclude with more than 80% of positive sentiments that can be seen from the high positive sentiment classification results in the labeling phase. Therefore, the use of social media X is suitable for expedition companies in Saudi Arabia to improve their skills.

References

Adak, A., Pradhan, B., Shukla, N. (2022). Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. *Foods*, *11*(10):1500. doi: 10.3390/foods11101500.

Al Sari, B., Alkhaldi, R., Alsaffar, D., Alkhaldi, T., Almaymuni, H., Alnaim, N., ... & Olatunji, S. O. (2022). Sentiment analysis for cruises in Saudi Arabia on social media platforms using machine learning algorithms. *Journal of big Data*, *9*(1), 21. https://doi.org/10.1186/s40537-022-00568-5.

Dang, N. C, Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, *9*(3), 483. https://doi.org/10.3390/electronics9030483.

Elhag, M., Abo, M., Idris, N., Mahmud, R., & Qazi, A. (2021). A multi-criteria approach for arabic dialect sentiment analysis for online reviews : exploiting optimal machine learning algorithm selection.*Sustainability*, *13*(18), 1-20.

Gulo, E.S., Gulo, Y.R., & Marbun, S.F. (2022). Comparison of the effectiveness of decision tree, naïve bayes, k-nearest neighbor and support vector machine algorithms in carrying out classification, *JUTIKOMP, 5*(2), 54-9.

Hakami, N.A. (2023). Identification of customers satisfaction with popular online shopping apps in Saudi Arabia using sentiment analysis and topic modeling. *In Proceedings of the 2023 7th International Conference on E-Commerce, E-Business and E-*

Government (ICEEG '23). Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3599609.3599610.

Mohammadi, M. Rashid, T. A., Karim, S. H. T., Aldalwie, A. H. M., Tho, Q. T., Bidaki, M., Rahmani, A. M., & Hosseinzadeh, M. (2021). A comprehensive survey and taxonomy of the SVM-based intrusion detection systems, *Journal of Network and Computer Applications*, *178*, 102983.

Mordor Intelligent Report, (2024). Saudi Arabia courier, express, and parcel (CEP) market trends. Available online: https://www.mordorintelligence.com/industry-reports/saudi-arabia-courier-express-and-parcel-cep-market. Last access: 1 May 2024.

Nisbet, R., Miner, G., and Yale, K, (2018). Handbook of Statistical Analysis and Data Mining Applications, Second Edition, Academic Press,

Prabhakar, E. Santhosh, M., Krishnan, A. H., Kumar, T. and Sudhakar, R. (2019). Sentiment analysis of US airline twitter data using new adaboost approach. *International Journal of Engineering Resource Technology*, *7*(01), 1–3, 2019.

Salem, M. A., & Md Nor, K. (2020). The effect of COVID-19 on consumer behaviour in Saudi Arabia: Switching from brick and mortar stores to E-commerce. *International Journal of Scientific & Technology Research*, *9*(07).

Semary, N. A., Ahmed, W., Amin K., Pławiak P., & Hammad M. (2024). Enhancing machine learning-based sentiment analysis through feature extraction techniques. *PLoS One*, *19*(2):e0294968. doi: 10.1371/journal.pone.0294968.

Sghaier, M. A., & Zrigui, M. (2016). Sentiment analysis for Arabic e-commerce websites. *In the proceedings of 2016 International Conference on Engineering & MIS (ICEMIS), Aga*dir, Morocco, 2016, pp. 1-7, doi: 10.1109/ICEMIS.2016.7745323.

Sobaih, A. E. E., & AlSaif, A. (2023). Effects of parcel delivery service on customer satisfaction in the Saudi Arabian logistics industry: Does the national culture make a difference? *Logistics*, *7*(4), 94. https://doi.org/10.3390/logistics7040094

Tuan, T. A., Long, H. V., Son, L. H., Kumar, R., Priyadarshini, I., & Son, N. T. K. (2020). Performance evaluation of Botnet DDoS attack detection using machine learning. *Evolutionary Intelligence*, *13*(2), 283-294.

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Neural Information Processing Systems*. https://arxiv.org/pdf/1706.03762.

Wang, N. Li, G., & Wang, Z. (2023). Fast SVM classifier for large-scale classification problems, *Information Sciences*, 642.

Wasiq, M., Johri, A., Singh, P. (2022). Factors affecting adoption and use of M-commerce services among the customers in Saudi Arabia, *Heliyon*, *8*(12), e12532. https://doi.org/10.1016/j.heliyon.2022.e12532.

Zhang, S., Zhong, H., Wei, C., Zhang, D. (2022) Research on logistics service assessment for smart city: A users' review sentiment analysis approach. *Electronics*, *11*(23), 4018. https://doi.org/10.3390/electronics11234018.

Zygiaris, S., Hameed, Z., Alsubaie, M. A., & Ur Rehman, S. (2022). Service quality and customer satisfaction in the post pandemic world: A study of Saudi auto care industry, *Frontiers in Psychology*, 2022, Sec. Organizational Psychology, 13 - 2022 https://doi.org/10.3389/fpsyg.2022.842141.