International Journal of Engineering Research & Management Technology
ISSN: 2348-4039
July-August 2023 Volume-10, Issue-4
Email: editor@ijermt.org www.ijermt.org

Experimental evaluation of various machine learning techniques for Cross Domain Sentiment Classification

Fahd Saleh Alotaibi*, Khaled Hamed Alyoubi#,
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
(E-mail: * fsalotaibi@kau.edu.sa, #kalyoubi@kau.edu.sa)

Abstract

Sentiment classification, a subfield of natural language processing, plays a vital role in understanding public opinions and sentiments towards various products, services, or events. However, sentiment classification models often struggle to generalize across different domains due to variations in language, vocabulary, and sentiment expression. To address this challenge, this research paper presents an experimental evaluation of various machine learning techniques for cross-domain sentiment classification. We compare and analyze the performance of several popular machine learning algorithms, including multiple variants of support vector machines (SVM), random forests, decision trees, Naïve bayes neural networks, in adapting sentiment classification models across different domains. In this study, we focussed on a diverse dataset Amazon Product review that consists of sentiment-labeled texts from multiple domains, such as product reviews, social media posts, and news articles. We then applied pre-processing techniques to handle domain-specific challenges, such as domain-specific jargon, slang, and abbreviations. Next, we implemented and trained different machine learning models using the collected data. Our experimental evaluation focuses on assessing the models' performance in terms of accuracy, precision, recall, and F1-score across various domains. The results of our experiments provide valuable insights into the strengths and limitations of different machine learning techniques for cross-domain sentiment classification. Furthermore, we identify the factors that contribute to effective cross-domain sentiment classification, such as feature selection, model complexity, and domain adaptation strategies.

Keywords: *Sentiment classification, machine learning, cross-domain, experimental evaluation, support vector machines, random forest.*

I. Introduction

Sentiment classification, a fundamental task in natural language processing (NLP), has gained significant attention in recent years due to its importance in understanding and analyzing public opinions and sentiments towards various products, services, or events. Accurate sentiment classification can provide valuable insights for businesses, marketers, and decision-makers, enabling them to gauge customer satisfaction, assess brand perception, and make informed strategic decisions. Sentiment analysis is used by various organisations for analyzing feedbacks of their products, movies, tweets etc. Traditional sentiment classification models primarily rely on supervised learning approaches, where they are trained on labeled data from a specific domain to classify sentiments in that domain accurately. However, these models often face challenges when applied to different domains. Variations in language use, vocabulary, and sentiment expression across domains can lead to decreased performance and limited generalizability of sentiment classifiers.

These techniques work efficiently but are not capable of achieving same levels of efficaciousness in the data which has more than one domain. Segregation of data into different domains is itself a task of classification resulting in higher degree of computation time and this paves the way for development of system that can work on multiple domains. Designing a system of classifiers to extract the useful features of interest of one domain and then to apply them for other domains is a useful technique. Cross-domain sentiment classification aims to overcome these limitations by developing models that can effectively classify sentiments across diverse domains. The goal is to create models that are capable of leveraging knowledge and patterns learned from one domain to improve classification performance in unseen domains.

In this research paper, we present an experimental evaluation of various machine learning techniques for cross-domain sentiment classification. We investigate the performance of popular machine learning algorithms, including Logistic Regression, Linear SVM, SVM-RBF, Decision trees, KNN, Naive Bias, in adapting sentiment classification models across different domains. The primary objectives of this study are twofold: firstly, to compare the performance of different machine learning techniques in cross-domain sentiment classification and identify the most effective approaches, and secondly, to explore the factors that contribute to successful cross-domain sentiment classification, such as feature selection, model complexity, and domain adaptation strategies.
To accomplish these objectives, we have used Amazon product reviews dataset. We have a total of 15 domains; each domain has positive and negative review. We have 1000 positive and negative for each product, therefore a total of 15000 positive and 15000 negative reviews. We pre-process the data to remove stop-words, handle domain-specific challenges, including domain-specific jargon, slang, and abbreviations. The experimental evaluation focuses on evaluating the performance of the machine learning models in terms of accuracy, precision, recall, and F1-score across various domains.
The remainder of this paper is organized as follows: Section 2 provides an overview of related work in cross-domain sentiment classification. Section 3 presents the methodology, including dataset collection. Section 4 discusses the experimental results and analysis. Finally, Section 5 concludes the paper, highlighting key findings and avenues for future research.

II. Literature Review

In the field of sentiment analysis, various techniques have been employed for analyzing sentiments across different domains. The existing literature highlights the significant utilization of several methods in this task. Traditional sentiment analysis has traditionally relied on manual feature selection [1]. Additionally, evolutionary approaches such as greedy heuristics [2] and genetic algorithms for feature reduction [3] have also been employed to enhance the performance of sentiment analysis. Genetic algorithms, in particular, have gained widespread popularity in studies related to feature reduction and optimization for sentiment analysis tasks [4][5][6][7]. However, it is important to note that sentiment analysis is often considered a domain-specific task. The effectiveness of traditional techniques is reliant on the availability of labeled data from each specific domain, making it computationally demanding. Consequently, traditional approaches may struggle to achieve comparable levels of accuracy when applied to data that encompasses multiple domains. The segregation of data into different domains itself poses a classification challenge, leading to increased computation time. Researchers are actively exploring new avenues for developing sentiment analysis systems that can effectively handle multiple domains. Recent studies have explored various approaches, such as random walk-based solutions [8], automatic extraction of domain features [9][10][11], and the use of multi-domain aspect-based methods [12][13][14][15], to address the challenges of sentiment analysis across multiple domains.
The emergence of deep learning and recurrent networks has also sparked significant research interest in sentiment analysis. Attention mechanisms, which extract important features associated with specific words, have shown promise in the field of natural language processing [16][17][18]. However, despite the availability of these works, several challenges persist in achieving effective sentiment classification across multiple domains. Many of these methods heavily rely on machine learning tools or neural networks for feature selection, which can introduce performance issues, encounter problems like the vanishing gradient, and necessitate a substantial amount of labeled data.While attention mechanisms offer a promising option for focusing on relevant features, their applicability and effectiveness in multiple domains still require further exploration and investigation. As sentiment analysis continues to evolve, researchers strive to overcome these challenges and develop robust techniques that can accurately classify sentiments in diverse domains.
One notable survey on cross-domain sentiment analysis is conducted by Li et al. (2018) titled "A Survey on Cross-

Domain Sentiment Analysis." The authors provide an extensive review of various approaches and techniques employed in cross-domain sentiment analysis. They categorize these techniques into four main groups: feature-based methods, transfer learning-based methods, domain adaptation methods, and ensemble methods. The survey paper discusses the advantages, limitations, and challenges associated with each category and provides insights into the current state of the field. Another survey by Zhang et al. (2019) titled "Cross-Domain Sentiment Analysis: A Survey" focuses on the challenges and techniques specific to sentiment analysis in different domains. The authors discuss domain adaptation techniques, transfer learning methods, and feature selection strategies used in cross-domain sentiment analysis. They also highlight the importance of domain knowledge and feature representation in achieving accurate sentiment classification across domains. In another survey by Zhou et al. (2020) titled "A Comprehensive Survey of Cross-Domain Sentiment Analysis," the authors provide an overview of both traditional and deep learning-based approaches for cross-domain sentiment analysis. They discuss various techniques, including feature-based methods, transfer learning, domain adaptation, and ensemble methods. The survey also addresses challenges such as domain discrepancy, data sparsity, and lack of labeled data in the target domain. Furthermore, Hu and Liu (2018) present a survey titled "Cross-Domain Sentiment Analysis: A Review." The authors have discussed different methods for cross-domain sentiment analysis, including domain adaptation, feature selection, and transfer learning. They also examine the use of linguistic resources, sentiment lexicons, and sentiment knowledge graphs in the context of cross-domain sentiment analysis. These surveys contribute to a better understanding of the research landscape, highlight the challenges faced, and provide guidance for future research in cross-domain sentiment analysis however none of these surveys provides experiment evaluation of various algorithms and are oriented towards some specific applications;

III. Proposed Methodology

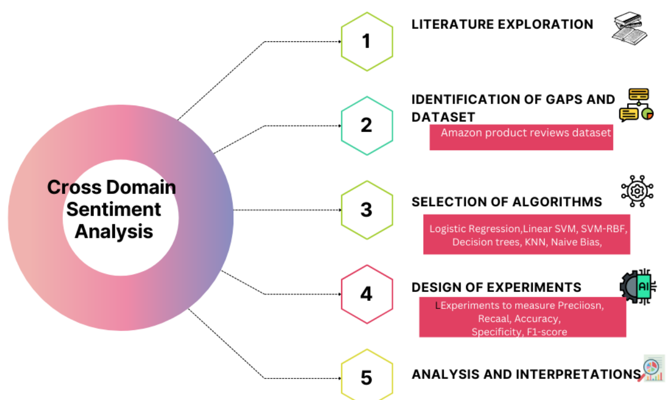

Fig 1:Methodology of the proposed work

The methodology adopted by this paper is shown in Figure 1. In the initial step of our research, we conducted a comprehensive exploration of the existing literature on cross-domain sentiment classification. This involved reviewing relevant research papers, survey articles, and related studies to gain a thorough understanding of the current state of the field. Based on our literature exploration, we identified certain gaps and research opportunities in the field of cross-domain sentiment classification. We identified that experimental evaluation of various algorithms for cross domain sentiment analysis is missing in literature. To address these gaps and conduct our experiments, we selected the Amazon product reviews dataset as our primary dataset. This dataset contains a wide range of reviews across different product categories, making it suitable for evaluating the performance of sentiment classification techniques in cross-domain scenarios.

To assess the performance of various machine learning techniques in cross-domain sentiment classification, we carefully selected a set of algorithms for our experiments. These algorithms include Logistic Regression, Linear SVM, SVM-RBF, Decision trees, KNN (K-Nearest Neighbors), and Naive Bayes. By choosing a diverse set of algorithms, we aimed to evaluate their respective strengths and weaknesses in handling sentiment classification across different domains. We designed a set of experiments to evaluate the selected machine learning algorithms' performance in cross-domain sentiment classification. Our experiments focused on measuring several key performance metrics, including Precision, Recall, Accuracy, Specificity, and F1-score. These metrics provided insights into the algorithms' ability to correctly classify sentiments in different domains and their overall effectiveness in sentiment classification tasks. Once the experiments were conducted and the results were obtained, we proceeded with the analysis and interpretation phase. We carefully analyzed the experimental data and compared the performance of the different algorithms across the various metrics. We also interpreted the results in the context of the specific dataset and identified patterns or trends that emerged from the experimental findings.

IV. Methods and Experiments

A. Logistic Regression

The first experiment was performed using Logistic Regression as a classifier to predict the product sentiment based on the reviews. Logistic Regression is a widely used model for modeling the probability of a specific class or event, such as positive/negative sentiment. It can also be extended to handle multiple classes. In our logistic regression model, we employed a logit function represented by Equation

Y(x) = b0 + b1x (1)

Here, Y(x) represents the predicted outcome, and x represents the input features. The parameters b0 and b1 are estimated through the training process to optimize the model's performance. To ensure that the output of our model falls within the range of 0 to 1, we applied the sigmoid function, as shown in Equation (2):

A(x) = 1 / (1 + e^(-x)) (2)

The evaluations of parameters such as accuracy, precision, and F1-Score are presented in Table 1. These results were obtained through the predictions made by the logistic regression model across various domains. These metrics provide insights into the model's performance in accurately classifying sentiments based on the reviews.

Domains	Accuracy	Precision	Recall	Specificity	F1-Score
DVD	0.7323	0.7516	0.6900	0.7700	0.7412
Books	0.7015	0.7159	0.6300	0.7500	0.6702
Kitchen	0.6750	0.6842	0.6500	0.7000	0.6667
Electronic	0.7326	0.7451	0.7000	0.7515	0.7416
Apparel	0.6375	0.6222	0.7000	0.5750	0.7416
Phones	0.6875	0.6667	0.7500	0.6250	0.7059
Computers	0.8250	0.7826	0.9000	0.7500	0.8372
Gourmet Food	0.7125	0.7073	0.7250	0.7000	0.7160
Grocery	0.6557	0.7500	0.7500	0.3151	0.7500
Music	0.7500	0.7500	0.7500	0.7500	0.7500
Musical Instruments	0.7050	0.7050	0.7050	0.7050	0.7050
Office Products	0.6875	0.6685	0.6758	0.6980	0.6721
Outdoors	0.6375	0.6503	0.6050	0.6800	0.6214
Software	0.6300	0.6368	0.6050	0.6550	0.6205
Sports	0.6350	0.6452	0.6000	0.6700	0.6218

TABLE 1. Results of logistic regression for cross domain sentiment analysis

The sentiment classification accuracy in computer domain is 82.50%, which is the highest among all the domains mentioned. Additionally, the Precision, Recall, Specificity, and F1-Score values for the Computers domain are also relatively high compared to other domains, with values of 78.26%, 90.00%, 75.00%, and 83.72%, respectively. These results indicate that the applied machine learning techniques achieved strong performance in accurately classifying sentiments within the Computers domain. The Precision, Recall, and F1-Score values for the Grocery domain were relatively lower compared to other domains, with values of 75.00%, 75.00%, and 75.00%, respectively. These results suggest that the applied machine learning techniques faced challenges in accurately classifying sentiments within the Grocery domain, potentially due to domain-specific complexities or limitations in the dataset

B. Linear SVMs

Linear SVM is a linear classifier, which means it assumes that the classes are linearly separable. In cases where the data is not linearly separable, additional techniques such as kernel methods can be used to transform the data into a higher-dimensional space, making it linearly separable. Linear SVM offers several advantages, including good generalization performance, ability to handle high-dimensional data, and resistance to overfitting. SVM aims to find a hyperplane that best separates the input data points into different classes. In the case of linear SVM, the algorithm seeks to find the optimal hyperplane that maximizes the margin between the closest points of different classes. The margin is the distance between the hyperplane and the nearest data points from each class, and the larger the margin, the more confident the model is in its classification. Mathematically, given a set of input vectors x and corresponding class labels y, linear SVM solves the following optimization problem:
minimize 1/2 ||w||^2
subject to y_i (w^T x_i + b) >= 1 for all training samples (x_i, y_i)
Here, w is the weight vector perpendicular to the hyperplane, b is the bias term, and ||w||^2 represents the Euclidean norm of the weight vector. The inequality constraint ensures that the data points are correctly classified with a margin of at least 1. The goal is to find the values of w and b that satisfy these constraints while minimizing the norm of w. The results of the conducted experiment are given in table 2.

Domains	Accuracy	Recall	Precision	Specificity	F1-Score
DVD	0.7112	0.7223	0.7456	0.7645	0.7248
Books	0.7183	0.7143	0.7143	0.7222	0.7143
Kitchen	0.6769	0.7143	0.6557	0.6410	0.6838
Electronic	0.7156	0.6522	0.7355	0.7613	0.6966
Apparel	0.6916	0.7105	0.7068	0.6923	0.7087
Phones	0.7551	0.7360	0.7632	0.7739	0.7494
Computers	0.7451	0.7221	0.7512	0.7623	0.7511
Gourmet Foods	0.7014	0.6872	0.7110	0.7159	0.6989
Grocery	0.7123	0.7015	0.7264	0.7155	0.7059
Music	0.7433	0.7511	0.7717	0.7340	0.7613
Musical Instruments	0.7563	0.7412	0.7845	0.7612	0.7741
Office Products	0.6846	0.6792	0.7143	0.6689	0.6964
Outdoors	0.6744	0.6689	0.7155	0.6656	0.6649
Software	0.6912	0.6894	0.7214	0.6845	0.7089
Sports	0.7112	0.7023	0.7165	0.6946	0.7055

TABLE 2. Results of linear svm for cross domain sentiment analysis

C. SVM-RBF

SVM-RBF is a non-linear extension of the linear Support Vector Machine (SVM) algorithm, allowing for more complex decision boundaries. The RBF kernel is a popular choice in SVMs as it can capture non-linear relationships between features. The RBF kernel measures the similarity or distance between pairs of data points in a high-dimensional space. It transforms the input features into a higher-dimensional space, where the data becomes more separable. SVM-RBF offers the advantage of being able to capture complex non-linear relationships in the data by mapping it to a higher-dimensional space using the RBF kernel. However, the selection of appropriate hyperparameters, such as gamma and C, is crucial for optimal performance. Improper tuning of these hyperparameters may result in overfitting or underfitting of the model.

Domains	Accuracy	Recall	Precision	Specificity	F1-score
DVD	0.7213	0.6915	0.7523	0.7626	0.7315
Books	0.6600	0.6500	0.6633	0.6700	0.6566
Kitchen	0.6750	0.6800	0.6733	0.6700	0.6766
Electronic	0.7313	0.7025	0.7538	0.7611	0.7269
Apparel	0.6216	0.6123	0.7012	0.6025	0.7212
Phones	0.6947	0.7157	0.6878	0.6735	0.7015
Computers	0.7412	0.7382	0.7540	0.7444	0.7460
Gourmet Food	0.7022	0.7027	0.7182	0.7018	0.7104
Grocery	0.7188	0.7059	0.7500	0.7333	0.7273
Music	0.7105	0.7000	0.7368	0.7222	0.7179
Musical Instruments	0.7005	0.7123	0.7345	0.7224	0.7178
Office Products	0.6845	0.6233	0.6522	0.6416	0.6626
Outdoors	0.6375	0.6503	0.6050	0.6080	0.6214
Software	0.6500	0.6565	0.6823	0.6232	0.6426
Sports	0.7005	0.7000	0.7268	0.7222	0.7179

TABLE 3 Results of SVM-RBF for cross domain sentiment analysis

D. Decision Trees

A decision tree is a modeling technique used for classification or regression tasks, represented in the form of a tree structure. It partitions a dataset into progressively smaller subsets while simultaneously constructing a decision tree. The resulting tree consists of decision nodes and leaf nodes. In the context of decision trees, two measures play a vital role: entropy and the Gini index.

Entropy quantifies the amount of information required to accurately describe a given sample. Mathematically, it can be expressed as:

Entropy = $-- \sum_{i=1}^{n} p_i * \log(p_i)$ (6)

Here, pi represents the probability associated with each class or domain. By calculating the entropy, we can evaluate the impurity or disorder present in the data.

On the other hand, the Gini index serves as a measure of inequality within a sample. It takes values between 0 and 1 and is calculated as the sum of the squared probabilities of each class:

Gini Index = 1- $\sum_{i=1}^{n} p_i^2$ (7)

The Gini index assesses the impurity of the data, where lower values indicate higher purity or homogeneity within the sample.

These measures, entropy and the Gini index, help guide the decision tree algorithm in determining the optimal splits and structure of the tree. By evaluating the impurity or inequality within the data, the decision tree algorithm can make informed decisions on how to partition the dataset, leading to more accurate classification or regression models.

Domains	Accuracy	Recall	Precision	Specificity	F1-Score
DVD	0.6012	0.5266	0.6154	0.6789	0.6212
Books	0.6600	0.6500	0.6633	0.6700	0.6566
Kitchen	0.6750	0.6800	0.6733	0.6700	0.6766
Electronic	0.5989	0.5165	0.6089	0.6716	0.5503
Apparel	0.6758	0.7228	0.7019	0.6173	0.7122
Phones	0.7611	0.7068	0.7468	0.6955	0.7656
Computers	0.7436	0.7077	0.7833	0.5990	0.7363
Gourmet Food	0.6565	0.5980	0.6662	0.6689	0.6263
Grocery	0.6235	0.6084	0.6264	0.6989	0.6566
Music	0.6798	0.7358	0.7027	0.6095	0.7189
Musical Instruments	0.6884	0.7156	0.7011	0.6621	0.7226
Office Products	0.6756	0.7388	0.7022	0.6166	0.6462
Outdoors	0.6642	0.7123	0.7114	0.6263	0.6466

Software	0.6923	0.7032	0.7205	0.6523	0.7033
Sports	0.7120	0.7231	0.7112	0.7003	0.7312

TABLE 4: Results of decision tree classifier for cross domain sentiment analysis

E. KNN

The k-nearest neighbors (kNN) algorithm is a straightforward yet powerful non-parametric method for classification. It operates by identifying the k nearest neighbors of a given data point, forming a neighborhood that assists in classification. However, the effectiveness of kNN hinges on selecting an appropriate value for k. The choice of k significantly impacts the accuracy of the classifier, as the method is sensitive to this parameter. Various approaches exist for determining the optimal value of k. One simple technique involves running the algorithm multiple times with different values of k and selecting the value that yields the best performance. This iterative process helps mitigate the bias introduced by the selection of k. It is important to note that the computational cost of kNN is relatively high since the classification occurs simultaneously for all training examples when they are first encountered. kNN has proven to be particularly effective in text-based classification tasks and has been widely utilized in such domains. In our model, we set k=7 as the chosen value. However, this value can be dynamically adjusted for different domains to identify the optimal k that yields the highest accuracy. Experimenting with different values of k enables us to adapt the model to the specific characteristics and complexities present in various domains..

Domains	Accuracy	Recall	Precision	Specificity	F1-Score
DVD	0.6415	0.5256	0.6478	0.7015	0.5859
Books	0.6612	0.6001	0.6451	0.6261	0.6714
Kitchen	0.6859	0.6550	0.6914	0.7159	0.6727
Electronic	0.6423	0.5254	0.6366	0.7015	0.5756
Apparel	0.6840	0.6509	0.6875	0.7159	0.6687
Phones	0.6955	0.6758	0.7110	0.7159	0.6930
Computers	0.6809	0.6842	0.6842	0.6774	0.6842
Gourmet Foods	0.7024	0.7333	0.6471	0.6774	0.6845
Grocery	0.7110	0.7500	0.6667	0.6774	0.7059
Music	0.7005	0.7222	0.6842	0.6774	0.7027
Musical Instruments	0.7067	0.7222	0.6842	0.6923	0.7027
Office Products	0.6456	0.6236	0.6641	0.6156	0.6542
Outdoors	0.6486	0.6667	0.6250	0.6319	0.6452
Software	0.6986	0.7297	0.6923	0.6667	0.7105
Sports	0.6974	0.7250	0.7073	0.6667	0.7160

TABLE 5: Results of knn for cross domain sentiment analysis

F. Multinomial Naive Bayes

Multinomial Naive Bayes is a popular algorithm for text classification tasks. It is specifically designed to handle discrete feature data, such as word frequencies or counts in documents. The algorithm assumes that features are conditionally independent given the class label, which is known as the "naive" assumption. In Multinomial Naive Bayes, the probabilities of feature occurrences in each class are estimated using the Multinomial distribution. During the training phase, the algorithm calculates the class priors (the probabilities of each class occurring in the training data) and the conditional probabilities of each feature occurring in each class. These probabilities are then stored in a model. When classifying new documents, Multinomial Naive Bayes calculates the likelihood of observing the given feature values (word frequencies) in each class. It combines this likelihood with the class priors using Bayes' theorem to calculate the posterior probability of each class. The class with the highest posterior probability is assigned as the predicted class label for the document. Multinomial Naive Bayes is widely used in various text classification applications, including sentiment analysis, spam detection, and topic categorization. It is known for its simplicity, computational efficiency, and ability to handle high-dimensional feature spaces commonly found in text data.

Domains	Accuracy	Recall	Precision	Specificity	F1-Score
DVD	0.7023	0.6586	0.7246	0.7556	0.6984
Books	0.6796	0.7326	0.5860	0.6411	0.6512
Kitchen	0.7108	0.7326	0.6811	0.7294	0.7059
Electronic	0.7012	0.6546	0.7213	0.7546	0.6923
Apparel	0.6934	0.7326	0.6117	0.6653	0.6667
Phones	0.7480	0.7326	0.7159	0.7608	0.7241
Computers	0.7429	0.7222	0.7222	0.7608	0.7222
Gourmet Foods	0.7307	0.8000	0.6993	0.6627	0.7436
Grocery	0.7704	0.8000	0.7692	0.7380	0.7843
Music	0.7546	0.7692	0.7692	0.7380	0.7691
Musical Instruments	0.7447	0.7692	0.7692	0.7143	0.7692
Office Products	0.7073	0.7000	0.7000	0.7143	0.7000
Outdoors	0.6985	0.7143	0.7143	0.6809	0.7143
Software	0.6907	0.7000	0.7000	0.6809	0.7000
Sports	0.7022	0.7209	0.7209	0.6809	0.7209

TABLE 6: Results of multinomial Naïve Bayes for cross domain sentiment analysis

G. Analysis and Interpretations

For the benefit of readers, we compared various models in terms of accuracy. The results are shown in Figure 2. The table provides accuracy values for different machine learning algorithms across various domains. Logistic Regression, SVM-RBF, and MNB (Multinomial Naive Bayes) consistently achieved relatively high accuracy across multiple domains, with accuracy values ranging from 0.6300 to 0.8250. Decision Trees had lower accuracy values compared to other algorithms, ranging from 0.5989 to 0.7611. SVM-Linear and KNN (K-Nearest Neighbors) showed varying levels of accuracy across domains, with accuracy values ranging from 0.6375 to 0.7563 for SVM-Linear and from 0.6409 to 0.7110 for KNN. The "Computers" domain consistently achieved the highest accuracy across multiple algorithms, while the "Software" and "Outdoors" domains generally had lower accuracy values.

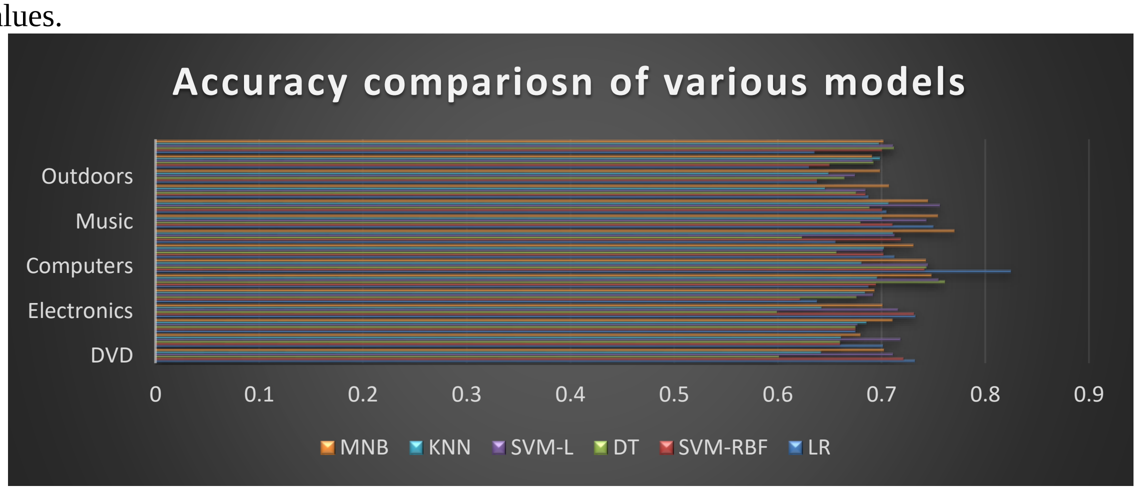

Fig 2: Accuracy comparison of each classifier for all the domains

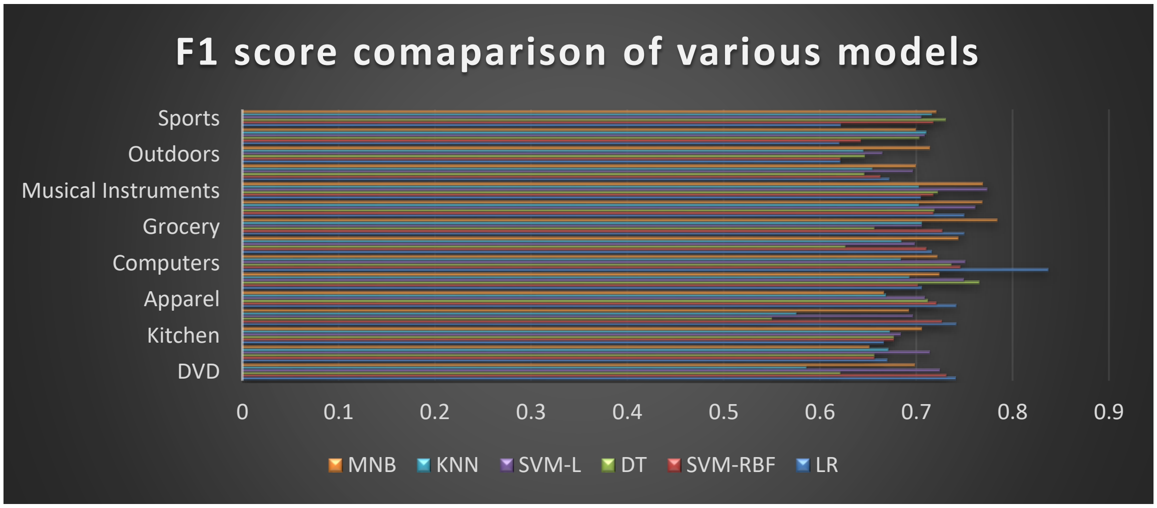

Fig 3: F1 score comparison of each classifier for all the domains

Figure 3 provides F1 scores for different machine learning algorithms across various domains. Logistic Regression (LR), SVM-RBF, and Multinomial Naive Bayes (MNB) achieved relatively high F1 scores across multiple domains. Decision Trees (DT) generally had lower F1 scores compared to other algorithms. SVM-Linear (SVM-L) and K-Nearest Neighbors (KNN) showed varying levels of F1 scores across domains. The "Computers" domain consistently achieved the highest F1 scores across multiple algorithms. The "Outdoors" and "Software" domains generally had lower F1 scores, indicating potential challenges in sentiment classification within these domains.

5. CONCLUSION

In this research paper, we conducted an experimental evaluation of various machine learning techniques for cross-domain sentiment classification. We explored the performance of different algorithms, including Logistic Regression, SVM-RBF, Decision Trees, SVM-Linear, KNN, and Multinomial Naive Bayes, across multiple domains. Based on the accuracy and F1 scores obtained from our experiments, we observed that Logistic Regression, SVM-RBF, and Multinomial Naive Bayes consistently achieved relatively high performance across domains. These algorithms demonstrated their effectiveness in accurately classifying sentiment in diverse domains, with the "Computers" domain consistently achieving the highest scores. On the other hand, Decision Trees generally exhibited lower performance compared to other algorithms, indicating potential limitations in handling cross-domain sentiment classification tasks. SVM-Linear and KNN showed varying levels of performance across domains, suggesting the need for further investigation and optimization in their application. Our findings highlight the importance of carefully selecting the appropriate algorithm for cross-domain sentiment classification. Logistic Regression, SVM-RBF, and Multinomial Naive Bayes emerged as promising choices for achieving accurate sentiment classification across diverse domains. Overall, this research contributes to the understanding of machine learning techniques for cross-domain sentiment classification and provides valuable insights for practitioners and researchers in selecting suitable algorithms for sentiment analysis tasks. Further research can focus on optimizing and enhancing the performance of these algorithms, as well as exploring other advanced techniques to improve cross-domain sentiment classification accuracy.

Acknowledgement

This research work was funded by Institutional Fund Projects under grant no. (G:1299-611-1440). Therefore, the authors gratefully acknowledge technical and financial support from Ministry of Education and Deanship of Scientific Research (DSR), King Abdulaziz University (KAU), Jeddah,
Saudi Arabia.

References

1. A. Duric, and F Song, "Feature selection for sentiment analysis based on content and syntax models" Decision support systems, vol 53(4), pp 704-711, 2012.
2. O. Gokalp, E. Tasci, and A. Ugur, "A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification". Expert Systems with Applications, vol 146, pp 113176, 2020.
3. F. Iqbal, JM Hashmi, B. C. Fung, R. Batool, A. M. Khattak, Aleem, S., and P.C. Hung, "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction". IEEE Access, vol. 7, pp 14637-14652, 2019.
4. S. Momani, Z. S. Abo-Hammour, , and Alsmadi, O. M. "Solution of inverse kinematics problem using genetic algorithms". Applied Mathematics & Information Sciences, vol. 10(1), pp 225, 2016
5. Abo-Hammour, Z., Arqub, O. A., Alsmadi, O., Momani, S.,and A. Alsaedi, "An optimization algorithm for solving systems of singular boundary value problems", Applied Mathematics & Information Sciences, vol. 8(6), article 2809, 2014
6. Abo-Hammour, Z., Arqub, O. A., Alsmadi, O., Momani, S.,and A. Alsaedi, "Optimization solution of Troesch's and Bratu's problems of ordinary type using novel continuous genetic algorithm". Discrete Dynamics in Nature and Society, vol. 9, article 2176, 2014.
7. O. Abu Arqub, Z. Abo-Hammour, Momani, S., and N Shawagfeh, "Solving singular two-point boundary value problems using continuous genetic algorithm" Journal of Abstract and applied analysis , Vol. 2012.
8. S. Xue, J. Lu, and G Zhang, "Cross-domain network representations", Pattern Recognition Journal , vol. 94, pp 135-148, 2016
9. J. Devlin, M. Chang, K Lee, K., and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint arXiv:1810.04805, 2018
10. Y. Du, He, M., Wang, L., and H. Zhang, "Wasserstein based transfer network for cross-domain sentiment classification" Knowledge-Based Systems, vol. 204, 106162, 2020
11. S. J. Pan, Ni, X., Sun, J. T., Yang, Q., and Z Chen, "Cross-domain sentiment classification via spectral feature alignment." In Proceedings of the 19th international conference on World wide web , 2014 pp. 751-760.
12. Poria, S., Cambria, E., and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network". Knowledge-Based Systems, vol 108, pp 42-49, 2016.
13. T. Rana, and Y. N Cheah, "A two-fold rule-based model for aspect extraction". Expert systems with applications, vol. 89, pp. 273-285, 2018
14. C. Wu, Wu, F., Wu, S., Yuan, Z., and Y Huang, "A hybrid unsupervised method for aspect term and opinion target extraction", Knowledge-Based Systems, vol. 148, pp. 66-73, 2018
15. H. Do, , Prasad, P. W. C., Maag, A., and A. Alsadoon. "Deep learning for aspect-based sentiment analysis: a comparative review", Expert systems with applications, vol. 118, pp. 272-299, 2019
16. Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, "Domain attention model for multi-domain sentiment classification" Knowledge-Based Systems, vol. 155, pp. 1-10, 2019
17. Liu, P., Qiu, X., and Huang, X. "Deep multi-task learning with shared memory". arXiv preprint arXiv:1609.07222.
18. Chauhan, G. S., Meena, Y. K., Gopalani, D., and R. Nahta, "A two-step hybrid unsupervised model with attention mechanism for aspect extraction", Expert systems with applications, vol. 161, article 113673, 2020