How to Cite:
Sivabharathi, G., & Chitra, K. (2022). Aspect-based sentiment analysis on current state of the market by classifying customer product reviews using support vector machine with several kernel learning. *International Journal of Health Sciences*, *6*(S1), 3168–3180.
https://doi.org/10.53730/ijhs.v6nS1.5371

Aspect-Based Sentiment Analysis on Current State of the Market by Classifying Customer Product Reviews Using Support Vector Machine with Several Kernel Learning

G. Sivabharathi
Part-Time Research Scholar, Madurai Kamaraj University, Madurai, Tamilnadu, India

Dr. K. Chitra
Assistant Professor, Government Arts College, Melur, Madurai, Tamilnadu, India

***Abstract*---**Aspect-based Sentiment Analysis identifies a customer's feelings about a given product feature. Aspect-based analysis is often used in product analytics to keep track of how a product is regarded by customers and what its strengths and weaknesses are. Product analytics is a subset of the classification problem. The Support Vector Machine (SVM) is a technology that can effectively tackle classification problems. In fact, SVM is a linear classifier, which implies it can only be used to classify data that is linearly separable. This approach should be enhanced with kernel learning to categorize non-linear separable data. This suggested system would create a classification model based on the Support Vector Machine, which will be updated with multiple kernel learning techniques and applied to a market data set for analytics. Sentiment Analysis examines a product's perception and market comprehension via the prism of sentiment data. To study customers, the proposed aspect-based sentiment analysis includes three levels of content analysis and four categories of opinions on a market data set. The results show that the proposed Several Kernel Learning with SVM has good accuracy and also has less computation time, where it needs about 61.46 seconds for training the data set and 23.26 seconds for testing the data set.

***Keywords*---**Market, Customer, Product.

Introduction

Every person has an attitude toward the things they encounter. Various folks can use the thing - there are dozens of such opinions from many persons. When all of these viewpoints are integrated, a distinct image of how the product is viewed emerges. Sentiment analysis can be beneficial to the business.

Definition: Sentiment Analysis

Sentiment analysis is a branch of Natural Language Processing that focuses on the investigation of subjective opinions or feelings about a topic gathered from multiple sources. It's a set of tools for identifying and extracting opinions and putting them to good use in the business.

Such algorithms sift through the text for information that reveals a user's attitude regarding the product in general or a specific feature. In other words, opinion mining and sentiment analysis provide an opportunity to investigate audience members' mindsets and examine the state of the product from a different perspective. As a result, sentiment analysis is an excellent tool for product and marketing evaluations.

Customer Sentiment Analysis

Customer Sentiment Analysis looks at how customers feel about a product and how they see the market via the prism of sentiment data. Sentiment analysis is a classification method that focuses on identifying an opinionated point of view and its disposition, as well as highlighting the information that is of particular interest in the process. There are numerous public and private data sources from which you can gain insight into the customer's impression of the product and the overall market position. User-generated product reviews, professional product reviews, and social media tractions are all examples of customer support contact for our product. Hence, Sentiment Analysis can be useful for the following operations as shown in the figure:

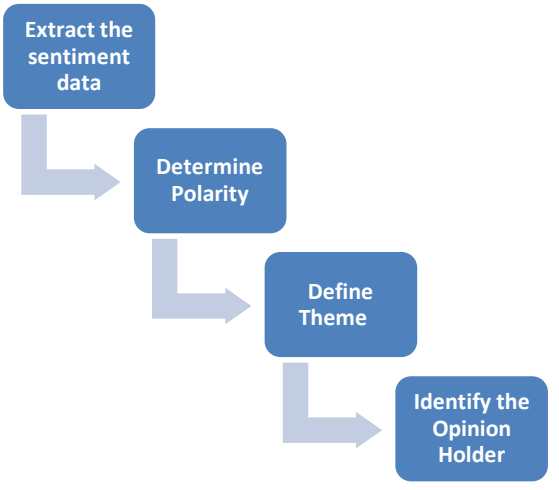

Figure 1. Operations of Sentiment Analysis

It begins by locating and extracting opinionated data on customer service and reviews. Then determine whether the polarity is positive or negative, as well as the topic matter (what is being discussed in general and specifically). Then locate the person who holds the viewpoint. Sentiment analysis algorithms can be employed in the following contexts, depending on the purpose:

- Document-level - this applies to the entire text.
- Sentence-level - gets a single sentence's sentiment.
- Sentiment of sub-expressions within a sentence is obtained at the sub-sentence level.

Mining an opinion is a difficult task due to its subjective nature. Different people have different viewpoints. Some things are worth more than others. "A view or judgement formed about something, not necessarily based on fact or understanding," according to sentiment analysis. An viewpoint is further divided into four subcategories:

- A direct opinion is one that expresses things directly. "The responsiveness of the menu in this website is bad," for example. You have a valid point here.
- A comparative opinion is one in which X and Y are compared based on a set of criteria. "The menu responsiveness in application Y is better than in application X," for example. It functions as micro-competitive research as well as providing insight into your product.
- Everything is fully described in the explicit opinion. "This wheel is moving," for example.
- Implicit opinions are those that are implied but not explicitly stated. "In two days, the app started lagging," for example. It is worth noting that implied viewpoints can include idioms and metaphors, further complicating the sentiment analysis process.

What it is not, i.e. polar message, can be used to determine neutral tone. Basically, you label anything that can't be classified as positive, bad, or a combination of the two as neutral. Sentiment analysis is the process of evaluating text using machine learning to determine the content's polarity. To perform sentiment analysis, researchers and data scientists employ a variety of text sources. Twitter, on the other hand, is a popular source for sentiment analysis. Twitter is regarded as one of the most important social media platforms. It serves as a podium for celebrities, athletes, politicians, field specialists, and scientists, among others, to express their views on a given topic. There are no restrictions on how you use Twitter. Users are free to create content that ranges from mundane activities to major incidents [5]. The importance of social media networks on our lives has grown to the point where social media is now used to gather first-hand knowledge on national and worldwide events [7].

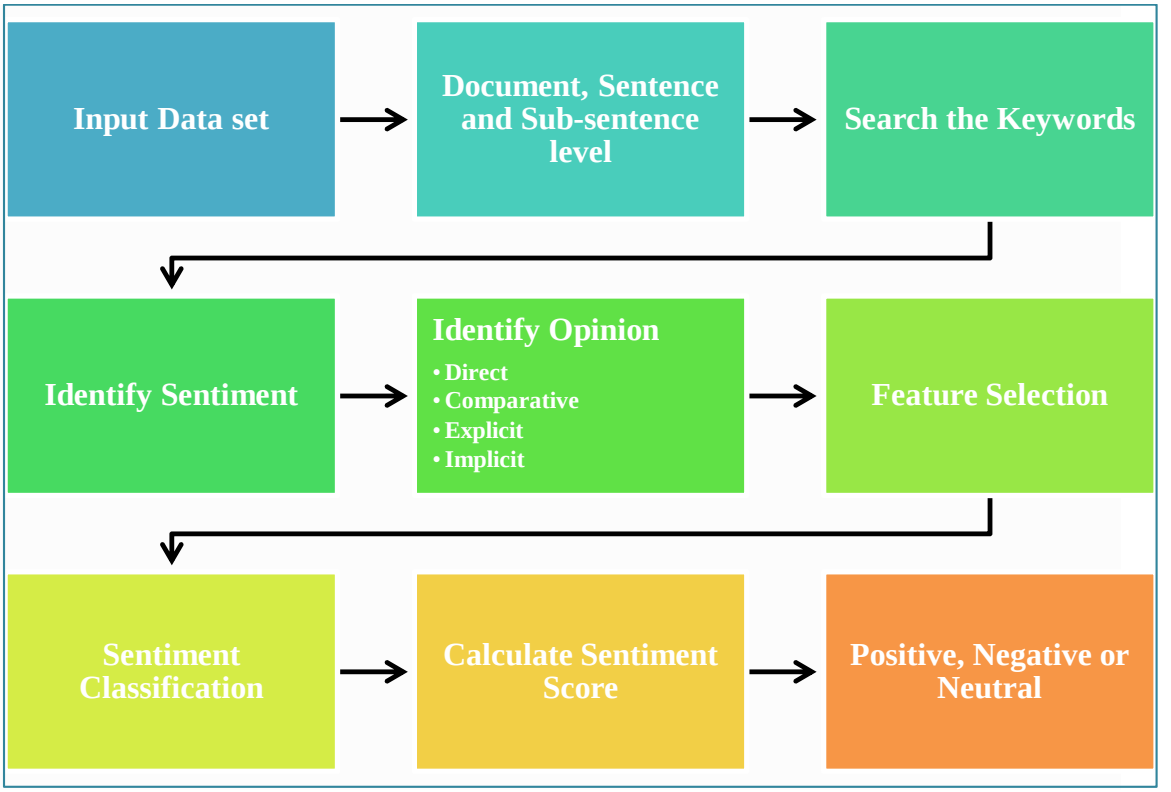

Figure 2. Process Flow in ASA-SVM

It is difficult to choose the right kernel throughout the learning process, hence several academics are working on a more flexible kernel learning system called Several Kernel Learning (SKL). In this study, we used a Support Vector Machine (SVM) with various kernel learning to classify the sentiment scores as positive, negative, or neutral.

Several Kernel Learning

The primary idea behind multiple kernel learning is to combine many kernel functions into a single function. For example, given some kernels K1(x, x′), K2(x, x′),..., KM(x, x′), the linear combination of the kernels can be represented as follows [3]:

$$K(x, x') = \sum_{m=1}^{M} d_m K_m(x, x') \quad (1)$$

Subject to

$$d_m \geq 0, \sum_{m=1}^{M} d_m = 1$$

Number M → Number of kernel functions.
If the best hyperplane for a single kernel can be found by solving the optimization issue below:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

Subject to

$$y_i(x_i.\ w + b) \geq 1 - \xi_i \forall i$$

$$\xi_i \geq 0 \qquad \forall i$$

where

C and ξ_i→Added variables for penalty value to reduce the number of misclassification data

In SKL, where decision function of the form

$$f(x) + b = \sum_{m=1}^{M} f_m(x) + b$$

$f_m$is associated with a kernel$K$.
So the optimization problem can be formulated as:

$$\min \frac{1}{2}\sum_{m=1}^{M} \frac{1}{d_m}\|f_m\|^2 + C\sum_{i=1}^{n} \xi_i \qquad (2)$$

Subject to

$$y_i \sum_{m=1}^{M} f_m(x_i) + y_i b \geq 1 - \xi_i, \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

$$\sum_{m=1}^{M} d_m = 1, d_m \geq \quad \forall m$$

Then, by using Lagrange function we get primal problem as:

$$\min L_p = \frac{1}{2}\sum_{m=1}^{M} \frac{1}{d_m}\|f_m\|^2 + C\sum_{i=1}^{n} \xi_i$$
$$+\sum_{i=1}^{n} \alpha_i \left(1 - y_i \sum_{m=1}^{M} f_m(x_i) - y_i b - \xi\right) - \sum_{i=1}^{n} \beta_i \xi_i$$
$$+\lambda\left(\sum_{m=1}^{M} d_m - 1\right) - \sum_{m=1}^{M} \mu_m d_m$$

$\alpha_i$andβ_i→Lagrange multiplier of constrains related to usual SVM problem
λandμ_m→constraints ond_m
Furthermore, if the coefficient dm vanishes and f(.) has to move to 0, the following fundamental problem can be given to the dual problem:

$$\max \sum_{i=1}^{n} \alpha_i - \lambda \tag{3}$$

Subject to

$$\sum_{i-1}^{n} \alpha_i y_i = 0$$
$$0 \le \alpha_i \le C, \forall i$$
$$\frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \le \lambda, \quad \forall m$$

Then, based on the Equation (2) and set value ofλso we get dual problem of Equation (3) as

$$\max_{\alpha} -\frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \sum_{m}^{M} d_m K_m (x_i, x_j) + \sum_{i=1}^{n} \alpha_i$$

Subject to

$$\sum_{i=1}^{n} \alpha_i \alpha_j = 0$$
$$0 \le \alpha_i \le C, \ \forall i$$

That dual problem is a linear constraint quadratic programming problem. Because training SVM is the same as solving a convex optimization problem, the SVM solution is both unique (provided K(xi, xj) is positive definite) and global optimal [4].

SKL-SVM: Several Kernel Learning with Support Vector Machine

One of the most important types of learning techniques used to identify data is distance learning. Two types of distance learning metrics were used in the proposed SKL-SVM: heterogeneous Euclidean overlap metric (H-EOM) and Euclidean distance. It uses attribute label information to calculate the distance between nominal characteristics. By minimizing the generalization error, SKL-SVM translates nominal features into actual space.

Because the space between numerical features is a normalized space, the suggested SKL-SVM method performs three levels of content analysis: DOCUMENT level, SENTENCE level, and SUB-SENTENCE level, and uses H-EOM to calculate the space concerned the ith nominal feature because the distance between numerical features is a normalized Euclidean distance, four sorts of

opinions can be identified: direct, comparative, explicit, and implicit. The greatest space, on the other hand, is the distance between unknown characteristics.

Experimental Setup

In this paper, the dataset is taken from the Amazon review data where Twitter data about Customer Product Reviews are easily available. This dataset contains:

Field Name	Meaning
reviewerID	ID of the reviewer e.g.B5AJSA233I5MMN
proID	ID of the product e.g.11151478
reviewerName	Name of the reviewer
useful	Usefulness rating of the review e.g.4/5
reviewerText	Text of the reviewer
rating	Overall rating of the product
outline	Outlineof the review
reviewTime	time of the review

In the experiment, sentiment analysis has been performed on this data using deep machine learning techniques like Neural Network and Linear Support Vector Machine and with the proposed SKL-SVM using the R platform. In this work, thedataset has been analyzed based on review text keywords like (Product, purchase, useful, good, bad, happy, etc) and classified the data into

- Positive (Strongly Positive, Weakly Positive),
- Negative (Strongly Negative, Weakly Negative)
- Neutral sentiment scores

The sentiment analysis is done over a different number of customer reviews on product and tried to define the sentiment scores linked with this data. Here, the analysis is carried out first on 25 customer reviews. After that, over 50 customer reviews have taken 150 reviews and, finally, over 500 reviews and classified them into different sentiment values categories.

The precision is measured as the ratio of the number of correctly identified Positive samples to the total number of Positive samples (either correctly or incorrectly). The precision of the model in categorizing a sample as positive is measured. The proposed SKL-SVM gives better precision than linear SVM and Neural Network.

$$Precision = \frac{True_{positive}}{True_{positive} + False_{positive}}$$

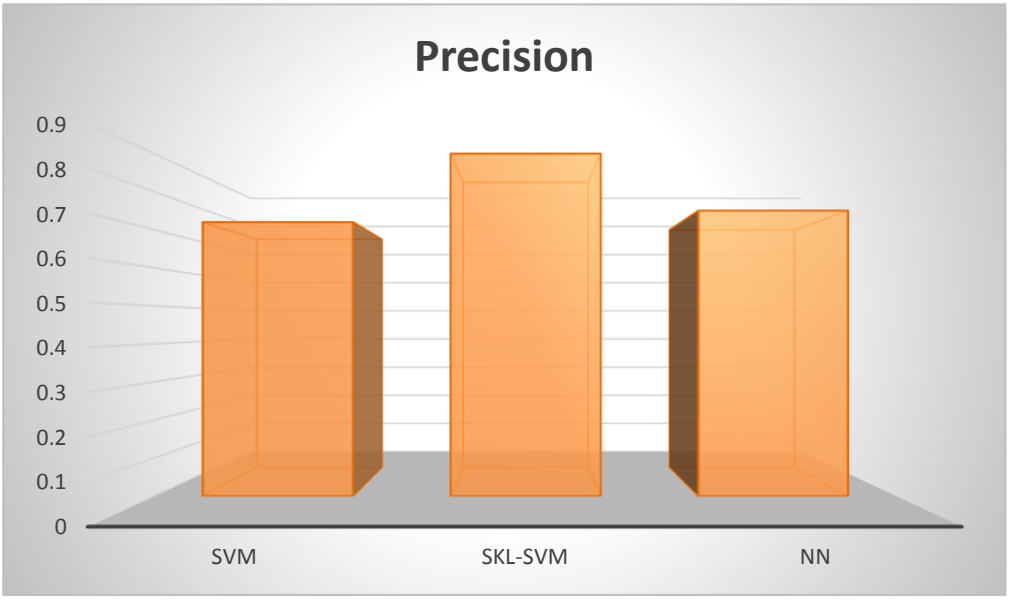

Figure 4. Precision

The recall is determined by dividing the total number of Positive samples by the number of Positive samples accurately categorized as Positive. The model's ability to recognize Positive samples is measured by the recall. The higher the recall, the greater the number of positive samples found. . The proposed SKL-SVM gives better recall than linear SVM and Neural Network

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}}$$

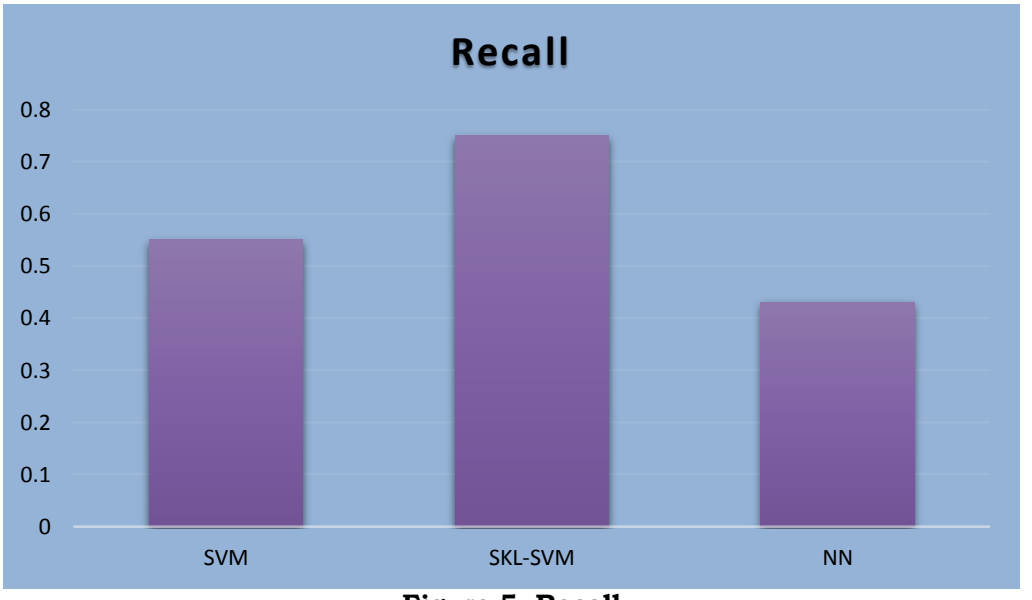

Figure 5. Recall

F1 Score

The weighted average of Precision and Recall is the F1 Score. As a result, this score considers both false positives and false negatives. Although it is not as intuitive as accuracy, F1 is frequently more useful than accuracy, especially if the class distribution is unequal. When false positives and false negatives have equivalent costs, accuracy works well. . The proposed SKL-SVM gives better F1 Score than linear SVM and Neural Network.

$$F1\ Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)}$$

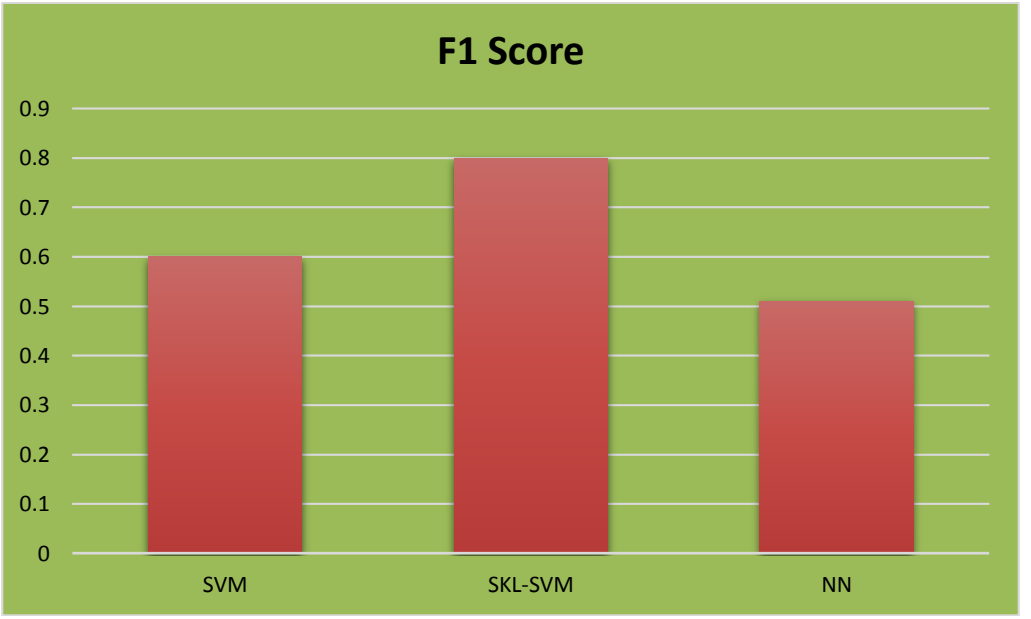

Figure 6. F1 Score

Accuracy is a statistic that sums up how well a model performs across all classes. It's helpful when all of the classes are equally important. The ratio between the number of right guesses and the total number of forecasts is used to compute it. The proposed SKL-SVM gives higher accuracy than linear SVM and Neural Network.

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

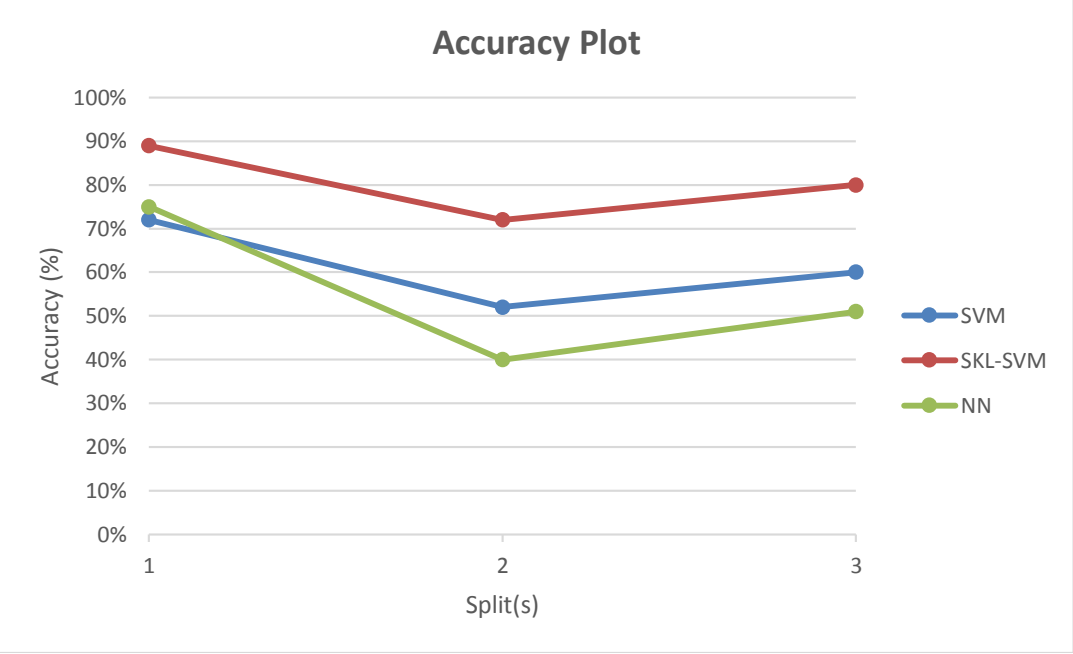

Figure 7. Accuracy Plot

Computation Time

The system was tested for computation time. The proposed SKL-SVM consumes about 61.46 seconds for training the data set and 23.26 seconds for testing the data set. Its computation time was compared with the Linear Support Vector Machine and Neural Network where it shows lesser training time and testing time than existing Linear SVM and Neural Network.

Figure 8. Training Time

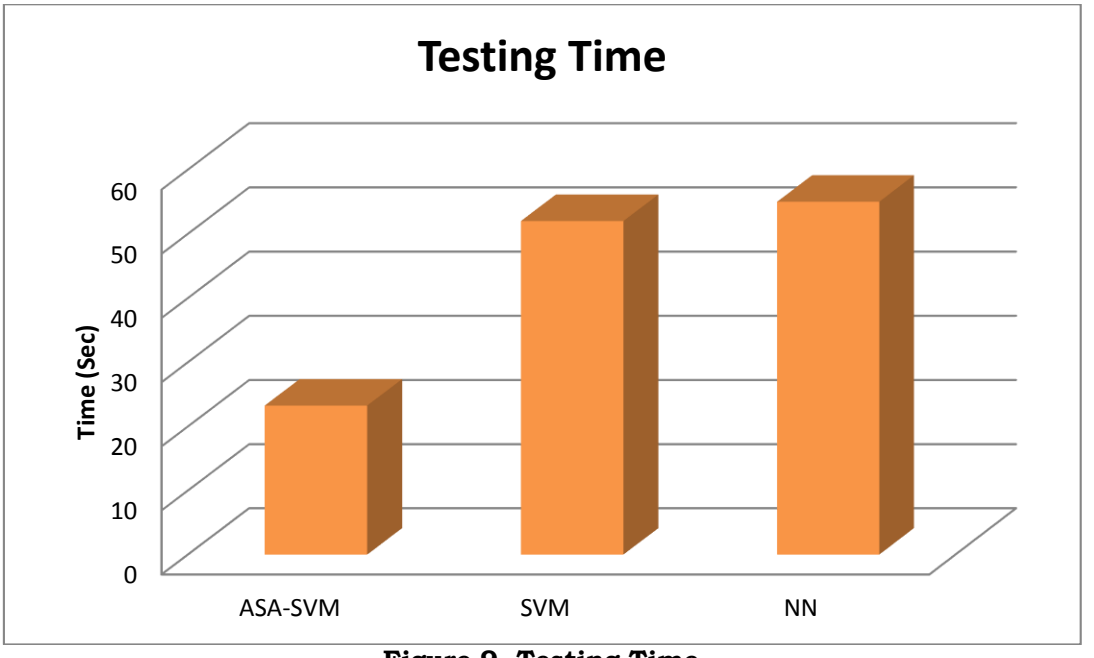

Figure 9. Testing Time

Conclusion

Sentiment analysis is a very useful tool for businesses since it allows you to receive honest feedback from your customers in an unbiased (or less biased) manner. When done correctly, it may offer a lot of value to your systems, apps, and web projects. The experimental results show that the proposed ASA-SVM which uses Support Vector Machine with Several Kernel Leaning has good accuracy and also has less computation time, where it needs about 61.46 seconds for training the data set and 23.26 seconds for testing the data set when it is compared with the Linear Support Vector Machine and Neural Network. Hence Overall, the proposed Aspect based Sentiment Analysis gives better viewpoint of the customers to analyze the market.

References

1) Berry, M.J.A., Linoff, G.S.: "Data Mining Techniques: for Marketing, Sales and Customer Relationship Management" (second edition), Hungry Minds Inc., 2004.
2) C. Rygielski, J. C. Wang, and D. C. Yeh, "Data mining techniques for customer relationship management," Technology in Society, vol. 24, 2002.
3) Chen, Y.-L., Tang, K., Shen, R.-J., Hu, Y.-H.:Market basket analysis in a multiple store environment", Journal of Decision Support Systems, 2004.
4) Cherian, J.; Gaikar, V.; Paul, R.; Pech, R. Corporate Culture and Its Impact on Employees' Attitude, Performance, Productivity, and Behavior: An Investigative Analysis from Selected Organizations of the United Arab Emirates (UAE). J. Open Innov. Technol. Mark. Complex. 2021, 7, 45. https://doi.org/10.3390/joitmc7010045
5) Chitra, K and et al, (2021), 'An impact of covid-19 on virtual learning: The innovative study on undergraduate students of Mumbai Metropolitan Region', Academy of Strategic Management Journal Volume 20, Special Issue 2, Pp. 1-19.
6) Humera Shaziya, G.Kavitha, Raniah Zaheer, 2015, Text Categorization of Movie Reviews for Sentiment Analysis, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue11.
7) *Jacob Cherian and* et al *(2020),* 'Impact of Web Content on Profitability and Market Share of Selected Newspapers: A Case Study of United Arab Emirates (UAE), International Journal of Management (IJM), Volume 11, Issue 10, October 2020, pp. 105-116. (Scopus Indexed Journal) http://www.iaeme.com/IJM/issues.asp?JType=IJM&VType=11&IType=10
8) Kamal, A. 2013 Subjectivity Classification using Machine Learning Techniques for Mining Feature-Opinion Pairs from Web Opinion Sources. International Journal of Computer Science Issues 10(5), 191- 200.
9) Redda, E.H., Gaikar Vilas, B., Tedla, B.A. (2021). E-Commerce Companies, Online Shopping and Customer's Satisfaction: A Comparative Study of Covid-19 Lockdown in India. Journal of Management Information and Decision Sciences, 24(S6), 1-17.
10) Sheikh, L Tanveer B. and Hamdani, "Interesting Measures for Mining Association Rules", IEEE Conference-INMIC, 2004.

11) Sivabharathi, G., Chitra, K., “Data Mining Techniques to Perform Market Analysis”, Journal of Emerging Technologies and Innovative Research (JETIR), Volume 7, Issue 3,2020.
12) Sivabharathi, G., Chitra, K., “Extracting Knowledge for Market Basket Analysis Using Association Rule”, Journal of Analysis and Computation (JAC), ISSN 0973-2861.
13) Sivabharathi, G., Chitra, K.,” Dynamic Euclidean K-Means Clustering Algorithm in Data Mining”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Vol.8, Issue-4, 2019
14) Sivabharathi, G., Chitra, K.,” Fuzzy C-Means Clustering Algorithm Using Initial Centroid Selection Process in Data Mining Techniques”, Turkish Journal of Computer and Mathematics Education, Vol.12 No.12,2021.
15) Sivabharathi, G., Chitra, K.,” Marketing Analysis Using Clustering in Data Mining” International Journal of Arts, Science and Humanities, Vol. 5 Special,Issue 1,2018.
16) Sivabharathi, G., Chitra, K.,” Sales Analysis of E-Commerce Using Data Mining Techniques”, Journal of Analysis and Computation (JAC), ISSN 0973-2861,2018.
17) Sivabharathi, G., Chitra, K.,” Survey on Data Mining Technique”, International Journal of Computer Engineering and Applications, Volume XII, Issue I, 2018.
18) Sonali Agarwal, Neera Singh, Dr. G.N. Pandey, “Implementation of Data Mining and Data Warehouse in E-Governance”, International Journal of Computer Applications, Vol.9, No.4, 2010.