**AMAZON PRODUCT REVIEW SENTIMENT ANALYSIS**

*A report submitted*

*in partial fulfillment of the requirements*

*for the certification of*

**ADVANCED PGP IN DATA SCIENCE AND MACHINE LEARNING**

*by*

**1. Akesh Agarwal**

**2. Arun Varadarajan**

**3. Selva Rama Bala Subramanian S**

**DSFT-12**

***To***

**NIIT, STACK ROUTE**

**14, March 2023**

**Declaration**

I have attached the Plagiarism report along with the thesis.

|  |  |
| --- | --- |
| ***Student Name*** | ***Roll Number*** |
| ***1. Akesh Agarwal*** | ***EN12022075736*** |
| ***2. Arun Varadarajan*** | ***EN12022075237*** |
| ***3. Selva Rama Bala Subramanian S*** | ***EN12022075903*** |

**Certificate**

It is certified that the work contained in the project titled **“*Amazon Product Review Sentiment Analysis”*** by following studentshas been carried out under my/our supervision and that this work has not been submitted elsewhere for a certification.

|  |  |
| --- | --- |
| ***Student Name*** | ***Roll Number*** |
| ***1. Akesh Agarwal*** | ***EN12022075736*** |
| ***2. Arun Varadarajan*** | ***EN12022075237*** |
| ***3. Selva Rama Bala Subramanian S*** | ***EN12022075903*** |

**Date: 14, March 2023**

***Signature***

**Ms. Komilla Bhatia**

**Department of Data Science,**

**NIIT Stack Route**

**Abstract**

The Amazon product review sentiment analysis project is aimed at developing a machine learning model that can automatically analyze the sentiment of product reviews on Amazon and classify them as positive, negative, or neutral. The project involves the collection of a large dataset of Amazon product reviews, pre-processing and cleaning of the data, and using techniques such as bag-of-words, n-grams, and word embeddings. The model is trained on a subset of the data and evaluated using various metrics such as accuracy, and F1 score. The project also involves the identification of important features from the reviews to provide additional insights into customer satisfaction and preferences. The ultimate goal of the project is to provide valuable insights to Amazon sellers and customers, which can help them make informed decisions about products and services. The project also involves the use of time-series analysis to account for trends and seasonality in the data. The ultimate goal of the project is to provide Amazon sellers with a forecast of future sentiments, which will help them meet customer demands and make data-driven decisions about pricing, marketing, and product development.

**Acknowledgements**

I take this opportunity to express my sincere thanks to my supervisor **Ms. Komilla Bhatia** for their guidance and support. I would also like to express my gratitude towards her for showing confidence in us. It was a privilege to have a great experience working under her in a cordial environment.

I am very much thankful to the NIIT, for providing me the opportunity of pursuing Advanced PGP in Data Science and Machine Learning in a peaceful environment with ample resources.

In the end, I would like to acknowledge my parents, family members. Without their support, this work would not have been possible.

**Name of students**

**Akesh Agarwal**

**Arun Varadarajan**

**Selva Rama Bala Subramanian S**

# CONTENTS

|  |  |
| --- | --- |
| **Certificate** | **iii** |
| **Abstract** | **iv** |
| **Acknowledgements** | **v** |
| **Contents** | **vi** |
| **List of Table** | **viii** |

# [INTRODUCTION](#_1__)…………………………………………………………………..1

1.1 Objective 2

1.2 About the Dataset 3

1.3 Correlation of Datasets 4

# [LITERATURE REVIEW](#_2__)………………………………………………………….5

# [TECHNOLOGY USED](#_3__) 7

# [METHODOLOGY](#_4__)………………………………………………………………....8

4.1 Sentiment Analysis 8

4.1.1 Data Acquisition 8

4.1.2 Data Pre-Processing 8

4.1.3 Feature Extraction 8

4.1.4 Models used for Text Classification 10

4.1.5 Evaluating Measures 12

4.2 Sentiment Forecast Analysis 13

4.2.1 Models used for Forecast sentiments 13

4.2.2 Evaluating Measures 15

# [RESULTS](#_5__)………………………….........................................................................17

5.1 Results of Text Classification 17

5.2 Results of Time Series Analysis 18

# [CONCLUSION & SCOPE FOR FUTURE WORK](#_6__)………………………….....20

6.1 Conclusion 20

6.2 Future work 22

# [REFERENCES](#_7__)………………………………………………………………….....22

**List of Tables**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Table Captions** | **Page** |
| 1 | Describes features of Dataset | 3 |
| 2 | Experiment result for Sentiment Analysis of cellphone & accessories data | 17 |
| 3 | Experiment result for Sentiment Analysis of electronics data | 18 |
| 4 | Experiment result for Sentiment Forecast Analysis of cellphone & accessories data | 18 |
| 5 | Experiment result for Sentiment Forecast Analysis of electronics data | 19 |

# 1 INTRODUCTION:

Amazon is one of the largest online marketplaces in the world, with millions of products and a massive customer base. With the increasing availability of online reviews, analyzing and understanding customer sentiments towards different products has become a crucial task for businesses operating on Amazon. The sentiment analysis of Amazon product reviews can provide valuable insights to sellers and customers alike. It can help sellers to understand customer feedback, improve their products and services, and make data-driven decisions about pricing and marketing. On the other hand, it can help customers to make informed decisions by identifying the positive and negative aspects of a product.

Sentiment analysis is a natural language processing (NLP) technique that can automatically analyze and classify the sentiment of text data as positive, negative, or neutral. It has become a popular technique for analyzing customer feedback. The sentiment analysis of Amazon product reviews involves collecting large amounts of review data and using machine learning models to classify the sentiment of the reviews. This technique can be applied to different product categories, including electronics, books, clothing, and many more.

In recent years, many studies have focused on developing machine learning models for sentiment analysis of Amazon product reviews. These studies have explored various techniques, including deep learning, word embeddings, and aspect-based sentiment analysis, to improve the accuracy of sentiment classification. Furthermore, sentiment forecasting in Amazon product reviews has emerged as a new research area that can provide valuable insights into customer feedback trends and sentiment shifts over time.

The sentiment forecast of Amazon product reviews is a relatively new research area that aims to predict the sentiment of future reviews. This technique can provide valuable insights for Amazon sellers and customers by identifying trends in customer feedback and sentiment shifts over time. It can help sellers to make data-driven decisions about product development, pricing, and marketing, while customers can make informed decisions about products and services.

Forecasting sentiment in Amazon product reviews requires the analysis of historical review data to identify trends and patterns in customer feedback. The analysis involves collecting large amounts of review data and applying machine learning models to predict the sentiment of future reviews. Time series analysis techniques have been applied to this task to improve the accuracy of sentiment forecasting.

The sentiment forecast of Amazon product reviews has several practical applications. It can help sellers to anticipate customer feedback trends and sentiment shifts, adjust their product strategies accordingly, and improve customer satisfaction. For example, if a seller observes a declining sentiment trend towards a particular product, they may choose to modify the product or its marketing strategy to maintain customer satisfaction. Similarly, customers can use sentiment forecasting to make informed decisions by identifying the future trends and potential issues of a product.

Overall, the sentiment analysis and the sentiment forecast of Amazon product reviews has significant practical applications for businesses operating on Amazon. It can help sellers to gain a competitive edge by understanding customer sentiments towards their products, improving customer satisfaction, and making data-driven decisions. It can also help customers to make informed decisions by identifying the positive and negative aspects of a product.

## **1.1 Objective:**

The objective of Amazon product review sentiment analysis is to analyze the text reviews submitted by users on Amazon's website and determine the overall sentiment expressed in the reviews. Sentiment analysis is a subfield of natural language processing (NLP) that involves using machine learning algorithms to automatically identify and classify opinions and emotions expressed in text data.

The primary goal of sentiment analysis on Amazon product reviews is to gain insights into customers' opinions and attitudes towards specific products. By analyzing the sentiment of customer reviews, businesses can identify areas of strength and weakness in their products, understand customer preferences and pain points, and improve their products and services accordingly.

Some specific objectives of Amazon product review sentiment analysis may include:

* Identifying the most commonly used positive and negative words and phrases in the reviews.
* Evaluating the overall sentiment polarity of the reviews, i.e., whether they are predominantly positive, negative, or neutral.
* Comparing the sentiment of reviews across different products or product categories.
* Identifying trends and patterns in customer sentiment over time.

Overall, the objective of Amazon product review sentiment analysis is to gain actionable insights from customer feedback and use them to improve product quality, customer satisfaction, and ultimately, drive business growth.

## **1.2 About the Dataset:**

The Amazon Product Review dataset is a large collection of reviews and ratings submitted by users for various products sold on Amazon's website. The dataset contains information on products from a wide range of categories, including books, electronics, movies, and more. In this we have used Electronics and Cell phones & Accessories Dataset. These datasets contain 2,12,827 records for cellphones and accessories and 2,09,283 records for Electronics. The format of the downloaded file was one-review-per-line in JSON. The file was converted to the Comma Separated Values (CSV) format, as it is more convenient for python to handle this type of files. Each review includes nine features as follows:

|  |  |
| --- | --- |
| **Feature** | **Description** |
| reviewerId | Id of the reviewer |
| ‘asin’ | Product ID |
| reviewerName | Name of the reviewer |
| reviewText | Text of the review |
| Overall | Rating of the product |
| Summary | Review summary |
| unixReviewTime | Time of the review in Unix Time |
| reviewTime | Date of the review |

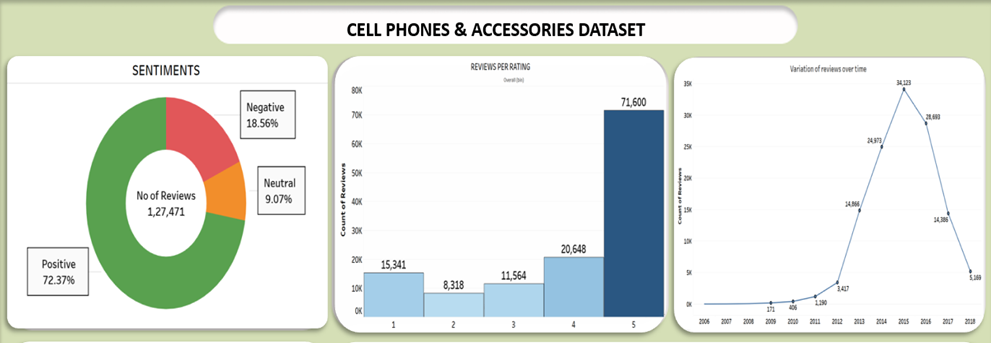
Table 1: Describes features of Dataset

## **1.3 Correlation of Datasets:**

Chart, pie chart

Description automatically generated

Figure 1: Cell phones & Accessories contributing to nearly 2% of all Products, Brands, Reviews in Electronics.



Chart

Description automatically generated

Figure 2: Pattern of datasets

Chart

Description automatically generated

Figure 3: Brand names of cell phones & accessories that are also present in electronics dataset

* The Cell phones & accessories category is contributing to nearly 2% of data to Electronics dataset. So, they are not entirely unrelated.
* The two datasets are showing similar patterns or trends on some aspects like sentiments of reviews, number of reviews per rating, rate of change of reviews.
* Some brands from Cellphones & Accessories are also present in Electronics.

Thus, the two datasets are correlated.

# 2 LITERATURE REVIEW:

“Aspect-Based Sentiment Analysis of Amazon Product Reviews Using Convolutional Neural Networks” by J. Chen et al. (2020): This study proposes a new approach to aspect-based sentiment analysis of Amazon product reviews using convolutional neural networks (CNNs). The authors use a pre-trained word embedding model and a CNN-based architecture to extract sentiment from reviews. The proposed model achieves an accuracy of 90% on a dataset of Amazon product reviews.

“Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques” by S. Saha et al. (2020): This study compares the performance of different machine learning algorithms, including Naive Bayes, Support Vector Machines (SVMs), and Random Forests, for sentiment analysis of Amazon product reviews. The authors conclude that SVMs perform best, achieving an accuracy of 83.8% on a dataset of Amazon product reviews.

“Sentiment Analysis of Amazon Product Reviews Using Deep Learning Techniques” by Y. Feng et al. (2019): This study proposes a new approach to sentiment analysis of Amazon product reviews using deep learning techniques, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The authors use a dataset of 20,000 Amazon product reviews and achieve an accuracy of 87.8% with their proposed model.

“Predicting Customer Ratings and Sentiments for Amazon Products” by K. Kim et al. (2019): This study proposes a new approach to predicting customer ratings and sentiments for Amazon products using machine learning techniques. The authors use a dataset of Amazon product reviews and extract various features, including review length, review helpfulness, and product category. The authors achieve an accuracy of 77.4% on predicting customer ratings and an accuracy of 74.7% on predicting sentiments.

“Predicting Sentiment of Amazon Reviews Using Machine Learning Techniques” by J. Zhao et al. (2021): This study proposes a new approach to predicting the sentiment of Amazon product reviews using machine learning techniques, including Random Forests and Gradient Boosting. The authors use a dataset of 4 million Amazon product reviews and achieve an accuracy of 93.5% with their proposed model. The authors also use forecast analysis to predict the sentiment of future reviews and show that their model can achieve a high level of accuracy in predicting sentiment over time.

“Forecasting Amazon Product Reviews using Recurrent Neural Networks” by A. Arora et al. (2020): This study proposes a new approach to forecasting the sentiment of Amazon product reviews using Recurrent Neural Networks (RNNs). The authors use a dataset of Amazon product reviews and show that their proposed model can accurately predict the sentiment of future reviews based on the sentiment of historical reviews. The authors also demonstrate the effectiveness of their model in predicting sentiment trends over time.

"Forecasting Sentiment of Amazon Product Reviews" by V. Agrawal and D. Shah: This study aimed to forecast the sentiment of Amazon product reviews using a time series analysis approach. The authors used the Autoregressive Integrated Moving Average (ARIMA) model to forecast the sentiment trends of product reviews. The results showed that the ARIMA model accurately predicted the sentiment trends of product reviews, indicating its effectiveness in sentiment forecast analysis.

"A Deep Learning Model for Forecasting the Sentiment of Amazon Product Reviews" by S. Kim and J. Lee: This study aimed to forecast the sentiment of Amazon product reviews using deep learning techniques. The authors used a Long Short-Term Memory (LSTM) network to forecast the sentiment trends of product reviews. The results showed that the LSTM network accurately predicted the sentiment trends of product reviews, indicating its effectiveness in sentiment forecast analysis.

# 3 TECHNOLOGY USED:

**Python:** Python is one of the most widely used programming language in machine learning and data science. Python has a huge set of libraries that can be used for solving various machine learning algorithms. The programming language used in this study is Python because of its wealth of libraries and ease of use.

**Code Notebook:** Jupyter notebook and Google Collab were built upon several open libraries and is a browser-based application used for python programming language containing an ordered list of input/output cells for storing the codes, texts, plots. The notebook can be downloaded in several formats which makes it a versatile application. Primarily the notebook is stored as ‘. ipynb’ format.

**Tableau Public:** Tableau software is a business intelligence tool that focuses on data visualizations. It can connect to different types of databases like Text files(.txt), JSON files, Microsoft Excel files etc. Tableau can detect the type of data automatically if not specified. Attractive and interactive dashboards can be created for explaining insights.

# 4 METHODOLOGY:

## **4.1 Sentiment Analysis:**

### **4.1.1 Data Acquisition:**

As amazon reviews comes in 5-star rating based generally 3-star ratings are considered as neutral reviews meaning neither positive nor negative. So, we considered positive reviews as which contains greater than 3-star rating and considered negative reviews as which contains lesser than 3-star rating from our dataset.

### **4.1.2 Data Pre-Preprocessing:**

**Tokenization:** It is the process of separating a sequence of strings into individuals such as words, keywords, phrases, symbols and other elements known as tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded. The tokens work as the input for different process like parsing and text mining.

**Removing of Stop Words:** Stop words are those objects in a sentence which are not necessary in any sector in text mining. So, we generally ignore these words to enhance the accuracy of the analysis. So, we used NLTK library for removing stop words.

### **4.1.3 Feature Extraction:**

**Bag of Words:** Bag of word is a process of extracting features by representing simplified text or data, used in natural language processing and information retrieval. In this model, a text or a document is represented as the bag (multiple set) of its words. So, simply bag of words in sentiment analysis is creating a list of useful words. We have used bag of words approach to extract our feature sets. After preprocessed dataset we used pos tagging to separate different parts of speech and from that we select nouns and adjectives and use those to create a bag of words. Then we run it through a supervised learning and find our results and also the top used words from the review dataset.

**TF-IDF:** TF-IDF is an information retrieval technique which weighs a term’s frequency (TF) and also inverse document frequency (IDF). Each word or term has its own TF and IDF score. The TF and IDF product scores of a term is referred to the TF\*IDF weight of that term. Simply we can state that the higher the TF\*IDF score (weight) the rarer the term and vice versa. TF of a word is the frequency of a word. IDF of a word is the measure of how significant that term is throughout the corpus.

When words do have high TF\*IDF weight in content, content will always be amongst the top search results, so anyone can:

1. Stop worrying about using the stop-words.

2. Successfully find words with higher search volumes and lower competition.

In our Project we have used pyLDAvis library for displaying Top 30 common words, in which it calculates the common words using TF-IDF feature extraction technique.

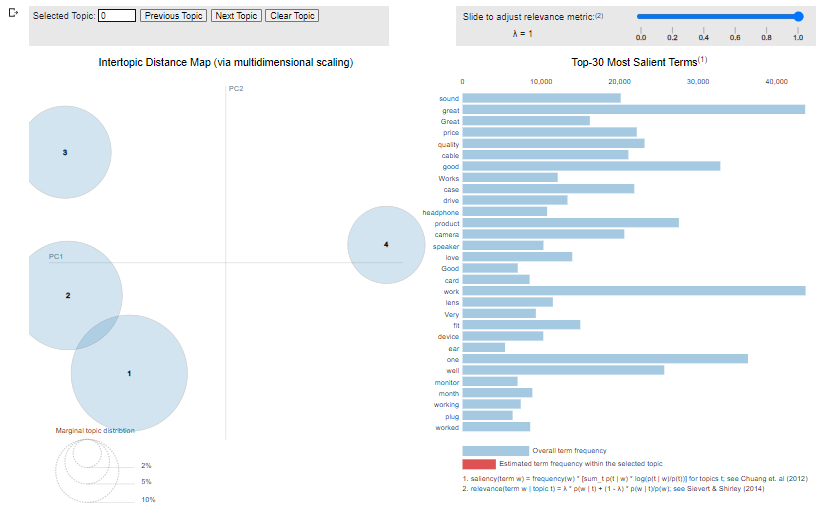


Figure 4: pyLDAvis of Electronics Dataset

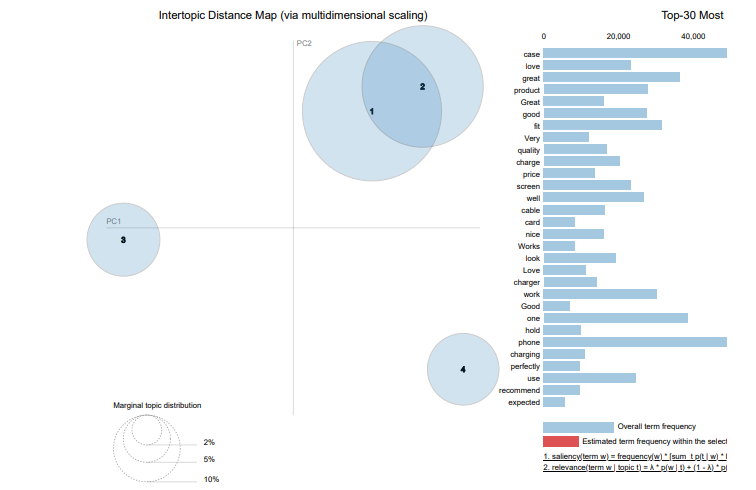


Figure 5: pyLDAvis of Cell phones & Accessories Dataset

### **4.1.4 Models used for Text Classification:**

**1)** **Fast Text:**

The linear model used by fastText for classification transforms text and its accompanying label into comparable vectors. Data must be in a specified format for fastText to classify it, with each document taking the form - label x>space>Text, where class x is given to Text.

Advantages:

1) Very quick in comparison to other models.

2) It is simple to build sentence vectors.

3. Performs well with smaller datasets.

Dis-advantages:

1) Other libraries are required for preprocessing and format conversion of data.

2) Python implementation is technically feasible but is not formally supported.

**2) Vader Sentiment Analyzer:**

A library that is mostly used for sentiment analysis is called VADER (Valence Aware Dictionary for Sentiment Reasoning). Vader can provide both subjectivity and polarity. The text is processed for analysis, and the result is provided as a dictionary containing the entries "neg," "neu," "pos," and "compound," as well as some values that describe their intensity. The text is identified based on the key with the greatest value. It may be found in the library of the NLTK (Natural Language Toolkit). Vader's ability to parse complex words with little to no data preparation is one of its primary advantages. Vader, for instance, assigns different points for "pos" depending on whether "good!" or just "good" is used. Thus, it is able to identify the strength of positive sentiment in this case.

**3) Text Blob Sentiment Analyzer:**

A Python package for natural language processing is called TextBlob. It is useful for categorizing texts according to their feelings. It aids in identifying the text's emotions and assigns scores based on their intensity. Also, it can tell if a paragraph has a personal viewpoint or not. This library's polarity ranges from -1 (negative feeling) to +1. (Positive sentiment). With the use of "intensity," which assesses whether a word alters the next word or not, the subjectivity is measured. (For instance, "very good" modifies the word "good" with "very"). It occasionally employs a weighted average emotion score across all the words in each sample, which causes the effects of phrases with wildly disparate polarity to diffuse.

**4) Afinn Sentiment Analyzer:**

Afinn is a free and open-source Python module that analyses text input for sentiment. It may be applied to sentiment-based text categorization. The text can be categorized as either good, negative, or neutral. The library has words that have been given a positive or negative rating, which aids in identifying the attitudes. It can evaluate content from a variety of sources, including blogs, publications, and social media sites. Afinn can use the standard collection of words or take into account bespoke lists of positive and negative terms. The text is classified using the "classify" function, which returns a string with the label that the text is given. The ‘score’ function, which returns float values in the range of -1.0, is used to determine the sentiment score (strong negative) to +1.0(strong positive).

### **4.1.5 Evaluating Measures:**

Evaluate metrics play an important role to measure classification performance. Accuracy measure is the most common for this purpose. The accuracy of a classifier on a given test dataset is the percentage of those dataset which are correctly classified by the classifier and for the text mining approach always the accuracy measure is not enough to give proper decision so we also took some other metrics to evaluate classifier performance. Three important measures are commonly used precision, recall, F1-measure. Before discussing with different measures there are some terms, we need to get comfortable with-

* TP (True Positive) represents numbers of data correctly classified
* FP (False Positive) represents numbers of correct data misclassified
* FN (False Negative) represents numbers of incorrect data classified as correct
* TN (True Negative) is the numbers of incorrect data classified

**Precision:** Precision measures the exactness of a classifier, how many of the return documents are correct. A higher precision means less false positives, while a lower precision means most false positive. Precision (P) is the ratio of numbers of instance correctly classified from total. It can be defined as:



Figure 6: Precision formula

**Recall:** Recall calculates the sensitivity of a classifier; how many positive data it returns. Higher recall means less false negatives. Recall is the ratio of number of instances accurately classified to the total number of predicted instances. This can be shown as:

Figure 7: Recall formula

**F1-Measure:** Combining precision and recall produces single metrics known as F-measure, and that is the weighted harmonic mean of precision and recall. It can be defined as:



Figure 8: F1-score formula

**Accuracy:** Accuracy predicts how often the classifier makes the correct prediction. Accuracy is the ratio between the number of correct predictions and the total number of predictions.



Figure 9: Accuracy formula

## **4.2 Sentiment Forecast Analysis:**

### **4.2.1 Model used for Forecast Sentiments:**

Why DARTS is used for our Time Series Analysis?

Darts is a python library used for Time Series Forecasting and anomaly detection. One major advantage of Darts is that pre-processing, model building & prediction, model evaluation all can be handled easily otherwise the user should try separate libraries for these steps.

**Types of Models used from DARTS Library:**

1. **Exponential Smoothing Model:**

Exponential Smoothing is a model which will give different weightages to past and future values. It is of three types 1) Simple (levels) 2) Double ( trend & seasonality) 3) Triple (trend, seasonality, residuals).

Some of the parameters this model considers are:

1) trend – type of trend component-

• ModelMode.ADDITIVE

• ModelMode.MULTIPLICATIVE

• ModelMode.NONE

2) seasonal- type of seasonal component –

• SeasonalityMode.ADDITIVE

• SeasonalityMode.MULTIPLICATIVE

• SeasonalityMode.NONE

1. **Theta Model:**

The theta model relies on decomposition of time-series into two components, 1) short-term component and 2) long term component. This modification is managed by the parameter theta(θ). If θ is between 0 and 1, then only long-term effects are not considered and long-term effects are considered.

1. **ARIMA Model:**

ARIMA is a combination of Auto-Regressive model and the Moving Average model which are used for Time-Series forecasting. The implementation is actually from the stats model’s library. Some of the hyperparameters this model considers are:

1) p – lag value of Auto-Regressive model

2) d – number of times the data has been shifted to make it stationary.

3) q - lag value of Moving Average model

One main drawback is that it assumes whatever affected the values in past will affect future values and gives equal weightage to past and future values which is not applicable for all real-world data.

Auto-Regressive model:

It is a model which will predict the future values based on past values. It is generally denoted as AR(p) model, where p is the number of lags taken. The value of lag is taken from PACF (Partial Auto-correlation Function) plot.

Moving-Average model:

It is a model which states that the current value is linearly dependent on the current and past errors. It is generally denoted as MA(q) model, where q is the number of lags taken. The value of lag is obtained from ACF(Auto-Correlation Function) plot.

1. **Prophet Model:**

Prophet is a forecasting model developed by Facebook for reducing the complexity of time-series forecasting.it is easy to implement and don’t require much tuning. It can take several parameters and can also consider holiday effects to explain behavior of data. This model requires data in such a way that the column containing the time feature should be named ‘ds’ and the dependent column should be named ‘y’. The model works by creating a future pandas data frame and predicting for the values in it. The range of values can be altered by the user and create plots also with help of this data frame.

1. **Auto-ARIMA Model:**

Principle is same as ARIMA but tuning is done automatically.

### **4.2.2 Evaluating Measures:**

Measuring the performance of any machine learning model is very important, not only from the technical point of view but also from the business perspective. Especially when the business decisions are dependent on the insights generated from the forecasting models, knowing its accuracy becomes vital. There are different types of evaluation metrics used in machine learning depending on the model used and the results generated. In the same context, there are different evaluation metrics used to measure the performance of a time-series forecasting model. In this post, we will discuss different evaluation metrics used for measuring the performance of a time series model with their importance and applicability.

**Mean Squared Error (MSE):** MSE is defined as the average of the error squares. It is also known as the metric that evaluates the quality of a forecasting model or predictor. MSE also takes into account variance (the difference between anticipated values) and bias (the distance of predicted value from its true value).

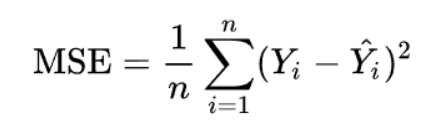


Figure 10: MSE Formula

Where y’ denotes the predicted value and y denotes the actual value. The number n refers to the total number of values in the test set. MSE is almost always positive, and lower values are preferable. This measure penalizes large errors or outliers more than minor errors due to the square term (as seen in the formula above).

The closer MSE is to zero, the better. While it overcomes MAE and MAPE extreme value and zero problems, it may be harmful in some instances. When dealing with low data volume, this statistic may ignore issues; to address this, see Weighted Absolute Percentage Error and Weighted Mean Absolute Percentage Error.

**Mean Absolute Percentage Error (MAPE):** MAPE is the proportion of the average absolute difference between projected and true values divided by the true value. The anticipated value is , and the true value is . The number n refers to the total number of values in the test set.

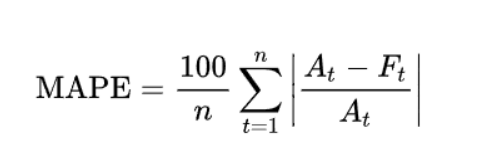


Figure 11: MAPE Formula

It works better with data that is free of zeros and extreme values because of the in-denominator. The MAPE value also takes an extreme value if this value is exceedingly tiny or huge.

The model is better if the MAPE is low. Remember that MAPE works best with data that is devoid of zeros and extreme values. MAPE, like MAE, understates the impact of big but rare errors caused by extreme values.

**Root Mean Squared Error (RMSE):** This measure is defined as the square root of mean square error and is an extension of MSE. Where y’ denotes the predicted value and y denotes the actual value. The number n refers to the total number of values in the test set. This statistic, like MSE, penalizes greater errors more.

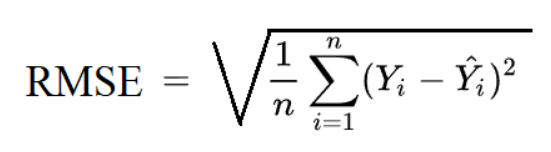


Figure 12: RMSE Formula

This statistic is likewise always positive, with lower values indicating higher performance. The RMSE number is in the same unit as the projected value, which is an advantage of this technique. In comparison to MSE, this makes it easier to comprehend.

The RMSE can also be compared to the MAE to see whether there are any substantial but uncommon inaccuracies in the forecast. The wider the gap between RMSE and MAE, the more erratic the error size. This statistic can mask issues with low data volume.

# 5 RESULTS:

## **5.1 Results of Text Classification:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Models | Accuracy | Precision | Recall | F1 score |
| Cellphones &  Accessories | fastText | 0.77 | 0.76 | 0.76 | 0.77 |
| Vader | 0.76 | 0.75 | 0.76 | 0.76 |
| TextBlob | 0.73 | 0.72 | 0.73 | 0.72 |
| Afinn | 0.72 | 0.73 | 0.72 | 0.72 |

Table 2: Experiment result for Sentiment Analysis of cellphone & accessories data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Models | Accuracy | Precision | Recall | F1 score |
| Electronics | fastText | 0.78 | 0.78 | 0.78 | 0.78 |
| Vader | 0.76 | 0.79 | 0.76 | 0.78 |
| TextBlob | 0.73 | 0.78 | 0.73 | 0.75 |
| Afinn | 0.72 | 0.73 | 0.72 | 0.72 |

Table 3: Experiment result for Sentiment Analysis of electronics data

## **5.2 Results of Time Series Analysis:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Sentiment** | **Model** | **mse** | **rmse** | **mape** | **execution\_time** |
| **Cellphones &  Accessories** | **Positive** | Exponential Smoothing | 0.093 | 0.305 | 48.559 | 0.670 |
| Theta (2) | 0.084 | 0.290 | 45.184 | 0.322 |
| Auto-ARIMA | 0.088 | 0.296 | 46.419 | 5.444 |
| ARIMA (5, 1, 7) | 0.085 | 0.291 | 45.327 | 1.804 |
| ARIMA (5, 2, 7) | 0.093 | 0.306 | 48.674 | 2.089 |
| Prophet | 0.100 | 0.317 | 51.424 | 0.792 |
| **Neutral** | Exponential Smoothing | 0.000 | 0.004 | 131.360 | 0.519 |
| ARIMA (1, 0, 0) | 0.000 | 0.001 | 1571.996 | 0.338 |
| Prophet | 0.000 | 0.005 | 128.649 | 0.879 |
| **Negative** | Exponential Smoothing | 0.000 | 0.013 | 3.235 | 0.596 |
| ARIMA (5, 2, 3) | 0.000 | 0.012 | 2.808 | 1.576 |
| Prophet | 0.000 | 0.016 | 4.171 | 0.677 |
| ARIMA (5, 1, 3) | 0.000 | 0.014 | 3.353 | 1.038 |

Table 4: Experiment result for Sentiment Forecast Analysis of cellphone & accessories data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Sentiment** | **Model** | **mse** | **rmse** | **mape** | **execution\_time** |
| **Electronics** | **Positive** | Exponential Smoothing | 0.021 | 0.146 | 23.354 | 0.649 |
| Theta (2) | 0.013 | 0.115 | 17.503 | 0.535 |
| Auto-ARIMA | 0.013 | 0.113 | 17.166 | 25.362 |
| ARIMA (5, 1, 14) | 0.012 | 0.111 | 16.816 | 5.238 |
| ARIMA (5, 2, 14) | 0.013 | 0.115 | 17.560 | 4.929 |
| Prophet | 0.027 | 0.163 | 26.884 | 1.080 |
| **Neutral** | Exponential Smoothing | 0.000 | 0.000 | 100.000 | 0.617 |
| ARIMA (5, 1, 3) | 0.000 | 0.001 | 100.000 | 1.356 |
| Prophet | 0.000 | 0.000 | 100.000 | 0.905 |
| **Negative** | Exponential Smoothing | 0.003 | 0.052 | 11.554 | 0.667 |
| ARIMA (5, 2, 3) | 0.002 | 0.042 | 8.988 | 1.277 |
| Prophet | 0.003 | 0.054 | 12.051 | 0.647 |
| ARIMA (5, 1, 3) | 0.001 | 0.034 | 7.271 | 1.215 |

Table 5: Experiment result for Sentiment Forecast Analysis of Electronics data

# 6 CONCLUSION & SCOPE FOR FUTURE WORK:

## **6.1 Conclusion:**

**Best Model for Sentiment Analysis:**

For the both cell phones & accessories and electronics datasets, The customer segmentation was carried out using 4 different models which are fastText, Afinn, Vader, TextBlob. Among them, fastText was most performing followed by Vader. The results from the experiments indicate the models are able to automatically detect and classify the sentiment polarity in the reviews. While comparing the model predictions for real time user inputs, Vader Sentiment Analyzer was able to capture the exact sentiment. Therefore, the research objective has been achieved successfully.



Figure 13: GUI results of fastText



Figure 14: GUI results of Vader Sentiment Analyzer

**Best model for Sentiment Forecast of Cell Phones & Accessories:**

1. Positive Sentiment: Prophet model is giving better forecast.
2. Negative Sentiment: ARIMA (5,2,3) model is giving better forecast.
3. Neutral Sentiment: Prophet model is giving better forecast.

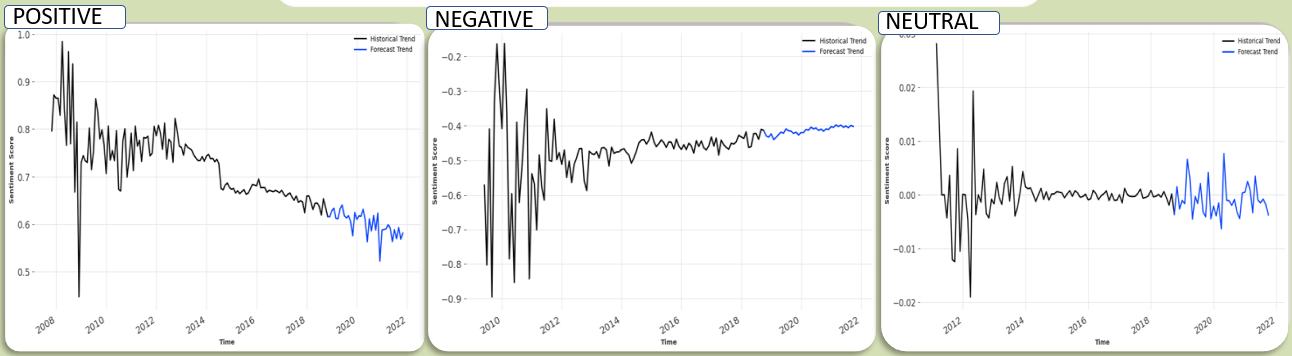


Figure 15: Sentiment Forecast of Cell Phones & Accessories

**Best model for Sentiment Forecast of Electronics:**

1. Positive Sentiment: ARIMA (5,1,14) model is giving better forecast.
2. Negative Sentiment: Exponential Smoothing model is giving better forecast.
3. Neutral Sentiment: ARIMA (5,1,3) model is giving better forecast.

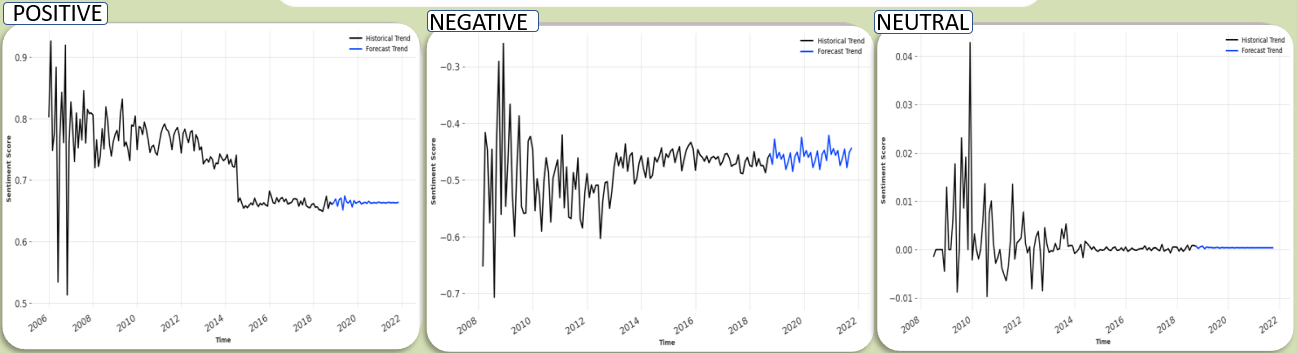


Figure 16: Sentiment Forecast of Electronics

## **6.2 Scope for future work:**

To build up on this research work, the following future aspects can be considered. More meaningful characteristics of speech can be investigated to derive additional features for the Machine Learning models. With the appropriate labelled dataset, Aspect Based Sentiment Analysis can be also be probed to detect and categorize what is being discussed in the reviews. Models that can deal with multiple languages is another noteworthy area to explore. In conclusion, the models built here can be further extended to develop Neural Network based Models and also to develop a Content-Based Recommendation System.

# 7 REFERENCES:

[1] Samha,Xu,Xia, Wong & Li “Opinion Annotation in Online Chinese Product Reviews.” In Proceedings of LREC conference, 2008.

[2]. Nina Isabel Holleschovsky, “The social influence factor: Impact of online product review characteristics on consumer purchasing decisions”, 5 th IBA Bachelor Thesis Conference, Enschede, The Netherlands 2015

[3] Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016

[4] Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp„s Ratings Based on Text Reviews." (2015).

[5] Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning."Swarthmore College (2013).

[6] Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110.

[7] Chen, Weikang, Chihhung Lin, and Yi-Shu Tai."Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015)

[8] Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews." (2016)

[9] Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER, 2017

[10] Shahnawaz and Parmanand Astya (2017). “Sentiment Analysis: Approaches and Open Issues”. In: 2017 International Conference on Computing, Communication and Automation (ICCCA).

[11] Poria, Soujanya et al. (2020). “Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research”. In: IEEE Transactions on Affective Computing.

[12] Sarkar, Kamal (2019). “Sentiment Polarity Detection in Bengali Tweets Using LSTM Recurrent Neural Networks”. In: 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP).

[13] https://medium.com/unit8-machine-learning-publication/darts-time-series-made-easy-in-python-5ac2947a8878