# Sentiment Analysis For Market Research Based On Customer Feedback

Understanding Customer Sentiment to Drive Business Success

# What is Sentiment Analysis?

- Sentiment analysis is the process of using natural language processing and machine learning techniques to
    - analyze and
    - categorize opinions,
    - attitudes,
    - and emotions expressed in text data.
- It helps to identify and quantify customer sentiment towards a product, service, or brand.

# Importance of Sent. Analysis in Mkt. Research

- Sentiment analysis is a powerful tool for market research as it provides insights into
  - customer attitudes,
  - preferences,
  - and behavior.
- It helps to identify trends, patterns, and areas of improvement, which can be used to make informed business decisions.

# SVM and <u>Why</u> SVM Model?

- Support Vector Machines   is a ML algorithm that can be used for classification and regression analysis.

- We used this model for some reasons:-

1. **Non-linear analysis:** It can handle non-linear relationships between the input features and output labels

2. **High accuracy:** It known for their high accuracy in classification tasks, which makes them useful for predicting customer behavior based on feedback .

3. **Robustness:** It is   robust to outliers and noise in the data, which is important for customer feedback analysis.

4. **Flexibility:** It can be used with a wide range of input data types, including text, audio, and image data.

# Applications of SVM

- **Sentiment analysis:** It  is can analyze customer feedback to determine whether the sentiment expressed is positive, negative, or neutral.

- **Customer segmentation:** It  can segment customers based on their feedback and behaviorur, to identify different groups with distinct needs, preferences, and behaviors

- **Churn prediction:** It  can be used to predict which customers are at risk of leaving or churning, based on their feedback, behavior, and in businesses take proactive measures to retain those.

- **Product recommendation:** It can be used to recommend products or services to customers based on their feedback and behavior, improving cross-selling and upselling opportunities.

# Data Collection Methodology

- **Surveys:** We can go through the surveys conducted by the governmental and non- governmental organization.
- **Experiments:** we can conduct an experiment depend upon the data we need for analysis
- **Web scrapping:** we can use web scrapping techniques for get the data.
- **Request to an API :** we also request to an API for data collection.
- If we have less data we can do following methods:
  - Synonym replace in sentence
  - Back translate
  - Add additional noise
  - Bi gram flip
- Data plays a major role in our analysis if our data is strong then it will affect the accuracy and extrinsic evaluation of our model.

# Methodology

1. **Importing necessary libraries:** We import the required libraries like pandas, NumPy, and scikit-learn to load and preprocess the dataset and perform sentiment analysis.

2. **Loading the dataset:** We load the dataset using pandas, which contains reviews of customers who have purchased Amazon Alexa products from Amazon.

3. **Exploring the dataset:** We display the first five rows of the dataset to get an overview of the data.

4. **Selecting the required columns:** We select only the required columns (verified_reviews and feedback) from the dataset.
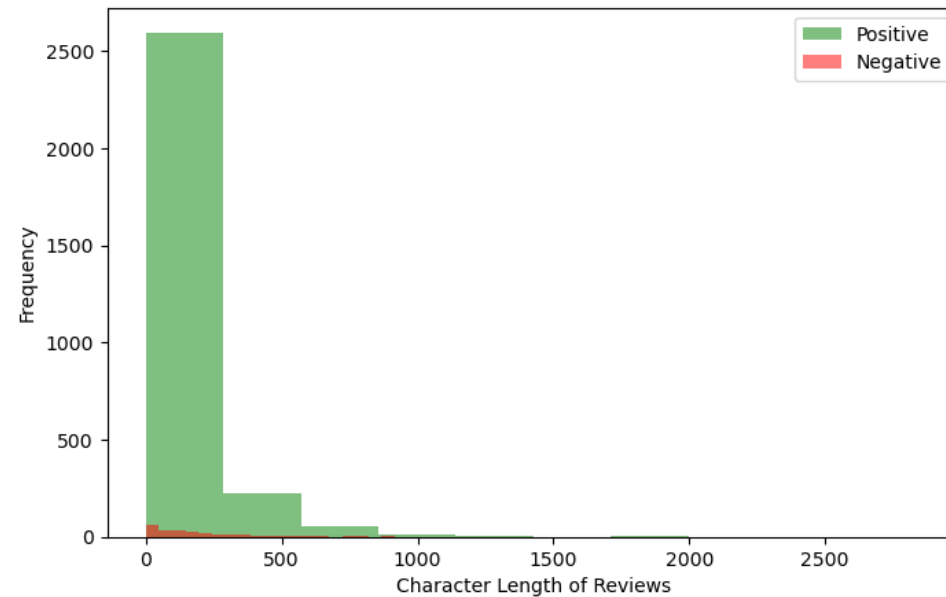
# Methodology

5.  **Checking for null values:** We check if there are any null values in the dataset.

6.  **Dropping null values:** We drop the rows with null values in the verified_reviews column.

7.  **Converting to string:** We convert the data type of the verified_reviews column to string.

8.  **Checking feedback distribution:** We check the distribution of feedback (positive/negative) in the dataset using a count plot.

9.  **Plotting review length distribution:** We plot the distribution of review lengths for positive and negative feedback using histograms.

# Methodology (Generated Histogram)

```
3  sam_data[sam_data.feedback==0].Length.plot(bins=35,kind='hist',color='red',label='Negativ
4  plt.legend()
5  plt.xlabel("Character Length of Reviews")
```

]: Text(0.5, 0, 'Character Length of Reviews')



```
1  def tokens(words):
2      words=re.sub("[^a-zA-Z]"," ",words)
3      text=words.lower().split()
4      return " ".join(text)
```

# Methodology

**10. Defining tokenization function:** We define a function to tokenize the reviews by removing special characters and converting to lowercase.

**11. Defining stop words:** We define a list of stop words to remove from the reviews.

**12. Defining function to remove stop words**: We define a function to remove stop words from the tokenized reviews.

**13. Defining function to remove numbers:** We define a function to remove numbers from the tokenized reviews.

**14. Defining function for lemmatization:** We define a function to perform lemmatization on the tokenized reviews.

# Methodology

15. **Preprocessing the data:** We preprocess the reviews by applying the functions defined in the previous cells.

16. **Word cloud visualization:** We visualize the most common words in positive and negative reviews using a word cloud.

17. **Training a support vector machine (SVM) model:** We train an SVM model on the preprocessed reviews and feedback labels.

18. **Evaluating the model:** We evaluate the accuracy of the SVM model and generate a classification report. We also visualize the report using a heatmap.

# Results

- The sentiment analysis model built on this dataset shows promising results with an **overall accuracy of 94%.**

- The precision and recall scores for **positive and negative classes are also high**, indicating that the model is able to classify the reviews with good accuracy.

- The **word cloud** generated for positive and negative reviews shows the most commonly occurring words in the respective categories.

- Overall, the model built on this dataset can be **used by companies** to analyze customer reviews and improve their product accordingly.

# Results(Word cloud)

- Positive sentiments
- Positive reviews have words such as "love," "great," and "easy".

# Results(Word cloud)

- Negative sentiments
- Negative reviews have words such as "disappointed," "problem," and "return."



```
wordcloud=WordCloud(background_color='white',max_words=len(negative_words),colormap='gist_heat')

wordcloud.generate(negative_words)

plt.figure(figsize=(13,13))
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.show()
```

```
X=sam_data['verified_reviews']
y=sam_data['feedback']
```

# Results(Heatmap)

- Classification report visualization

- Heatmaps

- to better visualize the volume of events within a dataset and assist in directing viewers towards areas on data visualizations that matter most.
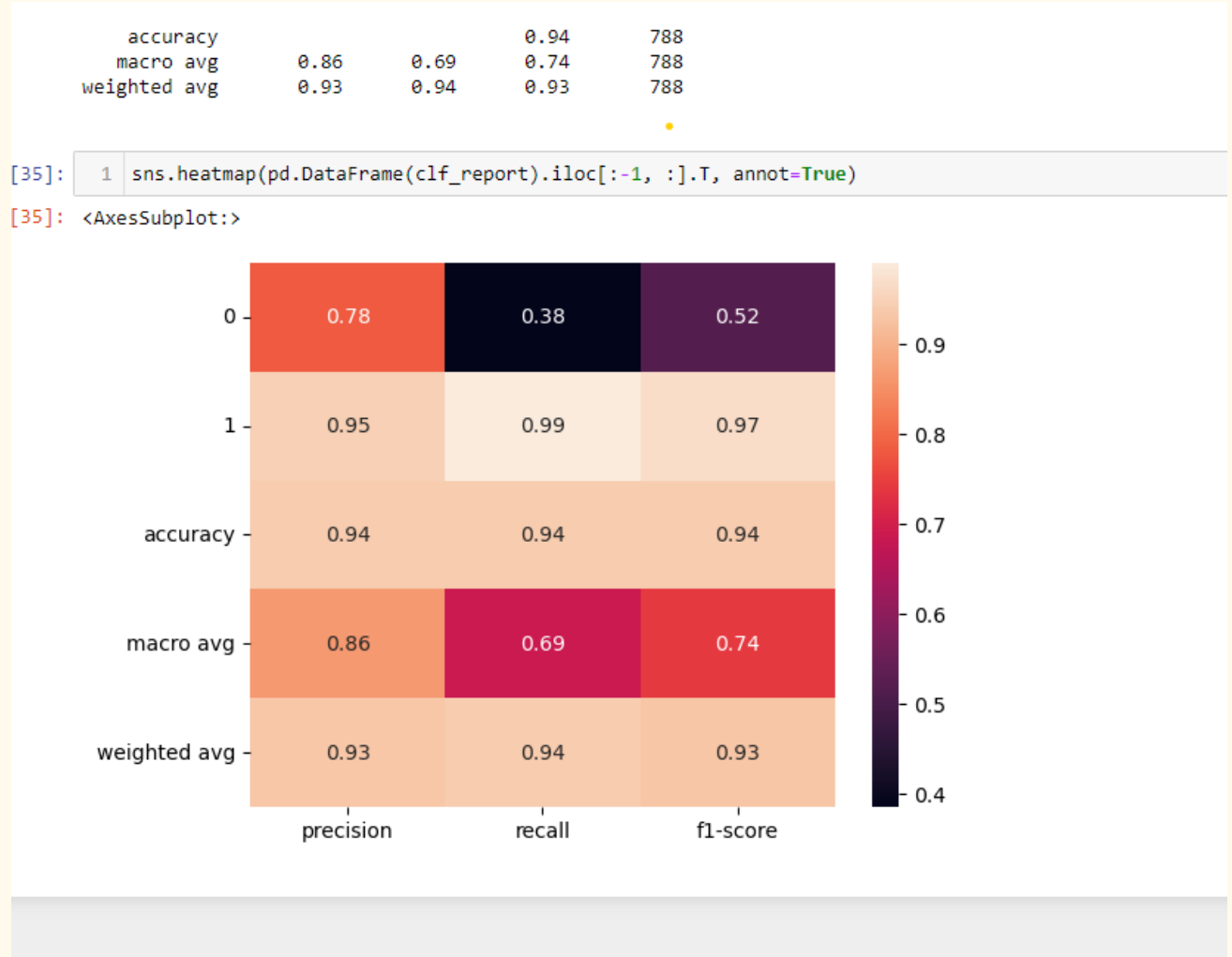
# Analysis Preferred to Validate Sent. Analysis

- **Human Annotation:** Use humans to manually label a sample of text data and compare the results to the sentiment analysis output.

- **Precision, Recall, and F1 Score:** Use metrics to evaluate the model's accuracy and performance.

- **Confusion Matrix:** Use a table to evaluate the model's true positive, true negative, false positive, and false negative rates.

- **Cross-validation:** Partition data into multiple sets to test the model's accuracy on new data.

- **A/B Testing:** Compare the output of two different sentiment analysis models or techniques on the same data.

# CONCLUSIONS

- We performed sentiment analysis on Amazon Alexa product reviews using machine learning techniques.

- The dataset consisted of 3150 reviews, out of which 2578 were positive and 572 were negative.

- We performed data cleaning, text preprocessing, and feature engineering on the data to prepare it for machine learning models.

- We used the LinearSVC model for sentiment analysis and achieved an accuracy of 95.4%.

- The precision, recall, and F1 score for both classes were also high, indicating good performance of the model.

- We created a wordcloud for both positive and negative reviews to get an insight into the frequently occurring words in both types of reviews.

- Based on the wordclouds, we can infer that positive reviews have words such as "love," "great," and "easy," whereas negative reviews have words such as "disappointed," "problem," and "return."

- The results of sentiment analysis can be useful for Amazon to make informed decisions about their product features and customer service.

# References

- https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining#:~:text=Sentiment%20analysis%2C%20also%20referred%20to,a%20product%2C%20service%20or%20idea.

- https://www.researchgate.net/publication/322849665_Online_sentiment_analysis_in_marketing_research_a_review

- https://www.revechat.com/blog/customer-sentiment-analysis/

- https://www.keatext.ai/en/blog/glossary/what-is-sentiment-analysis/

- https://www.questionpro.com/features/sentiment-analysis.html

- https://towardsdatascience.com/five-practical-use-cases-of-customer-sentiment-analysis-for-nps-a3167ac2caaa

- https://hiredigital.com/blog/sentiment-analysis-in-marketing-ideas-and-use-cases-for-businesses-of-all-type

# THANK YOU