ASU W. P. CAREY SCHOOL of BUSINESS
ARIZONA STATE UNIVERSITY

W. P. CAREY SCHOOL OF BUSINESS HONOR CODE MASTER'S PROGRAMS

We, the students of the W.P. Carey Master's programs, are committed to maintaining the highest levels of honesty, integrity, and professional courtesy for ourselves and amongst our classmates. We believe that ethical behavior is not only the cornerstone of a good academic program, but of good business as well. As such, we are steadfast in our commitment to the following principles:

- Academic Integrity – We view the W.P. Carey Academic Integrity Policy as a living, breathing document, not just concepts on a page. As such, we are resolute in ensuring that we abide by the principles contained within, both in letter and in spirit. Furthermore, we will hold our classmates to the same high standards to which we hold ourselves.

- Consideration of Others – As W. P. Carey students, we believe that we should treat others as we wish to be treated. To this end, we will be considerate of classmates, professors, and recruiters in all aspects. Furthermore, we will constructively add to the collaborative environment by helping and encouraging others.

- Professionalism – As W. P. Carey students, we believe that the behaviors we practice today will become habit in the future. To that end, we are committed to act in a professional manner in all school settings, including classes, networking events, and interviews. We will be cognizant of the fact that we are representing the program and our classmates at all times.

Full Name (printed): Amit Manocha

Signature ______ Date 5/5/2023

Full Name (printed): Helen Kim

Signature ______ Date 5/5/2023

Full Name (printed): Siri Kademani

Signature ______ Date 5/5/2023

Full Name (printed): Supratim Roy Chowdhury

Signature ______ Date 5/5/2023

Onsite MSBA Applied Project Report
Spring 2017
W.P. Carey, ASU

Topic	*Capturing customer sentiment by classification of Net Promoter Score (NPS) categories.*
Team	*23692*
Team Members	*Amit Manocha,* *Helen Kim,* *Siri Kademani,* *Supratim Roy Chowdhury*
Client information	*Axon Enterprise Inc., Customer Experience, Chris Nielsen, VP at Axon,* cnielsen@axon.com

Executive Summary– Supratim Roy Chowdhury

- **Background**: Axon Enterprise, Inc., is a prominent provider of public safety technology solutions that prioritizes customer experience, utilizing Net Promoter Score (NPS) as a top-level metric to measure customer satisfaction and to incentivize all Axon employees to prioritize customer needs.
- **Goal & Objective**: The goal of the project is to improve Axon's NPS score by better understanding customer feedback. This will be achieved by developing a machine learning model to categorize free-text responses and generating reports to identify areas for improvement. Ultimately, the objective is to assist Axon enhance the customer experience and drive business growth.
- **Methodology**: The methodology involved creating 6 binary classification models for each of the categories using Spacy to categorize texts. Next, the models were used to classify customer feedback into the respective categories. The feedback was then analyzed to identify areas of improvement for enhancing customer experience. Finally, the results were reported to stakeholders for further action.
- **Deliverables**: The deliverables include self-improving machine learning (ML) models that can classify customer feedback into the relevant categories, and Tableau dashboards that provide insight on areas of improvement for customer experience.
- **Future Scope**: Automating the categorization process by hosting the model on a server, enhancing the model to predict customer dissatisfaction, and regularly updating the model with new feedback to improve efficiency. Additionally, a web application could be developed for real-time monitoring of customer satisfaction based on primary categories.

Background – Helen Kim

Axon Enterprise, Inc. is the leading provider of public safety technology solutions, which includes body cameras and digital evidence management systems. The company values its relationship with customers, and therefore is deeply invested in opportunities to enhance customer experience. To measure the quality of customer experience, Axon runs a customer satisfaction survey regularly throughout the year, gathering thousands of responses annually. In addition to the qualitative feedback, the survey also asks customers to rate the company on a scale of 1 to 10 called the Net Promoter Score (NPS). NPS is *the* key metric as it not only measures customer satisfaction but also is a measure to incentivize Axon employees to prioritize customer needs. The obtained feedback from the surveys, along with the corresponding NPS scores, are then manually classified into 6 categories to be analyzed to identify potential areas of improvement to ultimately enhance the customer experience.

The 6 categories of customer feedback are defined as follows:

- Product: Related to specific products, or functions/features of products.
- Cost: Cost-related responses from the customers post or pre-purchase.
- Sales: Pertains to the sale of the products, the product purchasing cycle, and the teams who are responsible for quote building, generating invoices, and contracts.
- Training: Pertains to training sessions that are provided for customers regarding the use of the product, and the certifications for product usage.
- Customer Service: Pertains to the post-purchase experience or the ongoing service experience with Axon.
- Others: General comments about Axon and those without the mention of any specific category.

The categories for NPS are as follows:

- Detractors: 1~6
- Passives: 7~8
- Promoters: 9~10

Problem Statement – Helen Kim

Axon continuously strives to enhance the customer experience by identifying problem areas through NPS and as well as various customer surveys and operational data sources to proactively address any potential issues. However, the current process of feedback categorization relies on the manual labor of a single staff, therefore, is time-consuming, inefficient and prone to human error. Manual analysis also lacks the ability to classify free-text feedback quickly and accurately into actionable categories. Especially with Axon growing every day and hence observing a increased number of customer feedback, the current process may lead to delayed response times and missed opportunities to enhance customer experience, which can ultimately result in lower NPS scores and negatively impact the business.

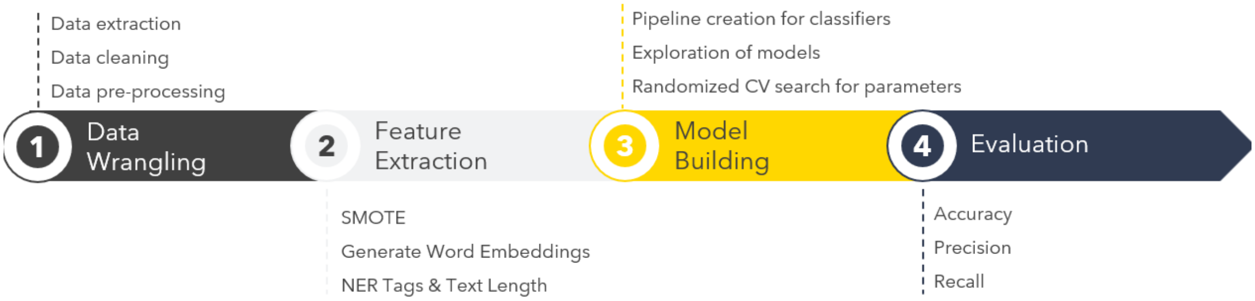
To address the issues with the existing process, the project aims to leverage a self-learning machine learning (ML) model that can analyze and categorize customer feedback through an automated process. Utilizing a ML model will increase accuracy and efficiency of categorization and help gain insight into the problem areas, which is expected to improve the NPS score and overall customer satisfaction for interactions with Axon.

In addition, a Tableau dashboard will be created to visually represent the categorized feedback data and provide insight into the key areas of improvement. This will assist Axon in analyzing feedback more efficiently, identifying areas for improvement and providing recommendations for enhancing their customer experience.

Methods – Amit Manocha

The methodology involved creating 6 binary classification models for each of the categories using Spacy to categorize texts. Next, the models were used to classify customer feedback into the respective categories. The feedback was then analyzed to identify areas of improvement for enhancing customer experience. Finally, the results were reported to stakeholders for further action.

This is the process flow we followed:



<u>Data Wrangling:</u>

Data Extraction:

We have used 2 datasets provided by Axon.

Data Cleaning:

- Manually aligning previous year data with the latest data and merged them together.
- Removing uncategorized data
- Concatenating various sub-categories into single primary category
- Combining ‘Customer Success’, ‘Customer Experience’, and ‘Professional Services’ under ‘Customer Service’

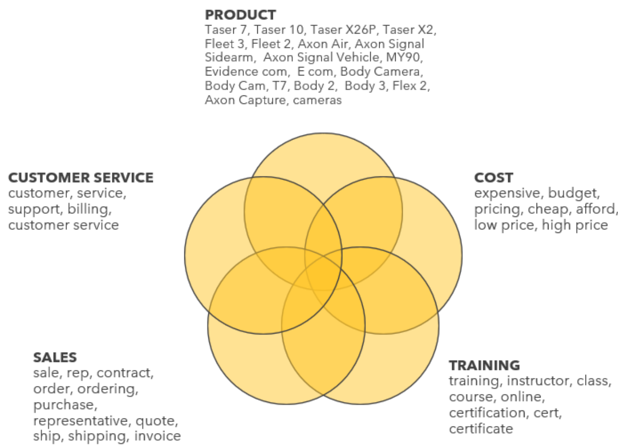
Data Pre-processing:

- Duplicate Removal
- Data was passed through the NLP Pipeline (en_core_web_lg)
    - Stop Words Removal
    - Lemmatization
    - Eliminate Punctuations
    - Conversion to Lower Case

<u>Feature Extractions:</u>

The below Axon product list as well as frequent words across all categories were input to Named Entity Recognition (NER) to add bias to the imbalanced class.

Additional Features generated:

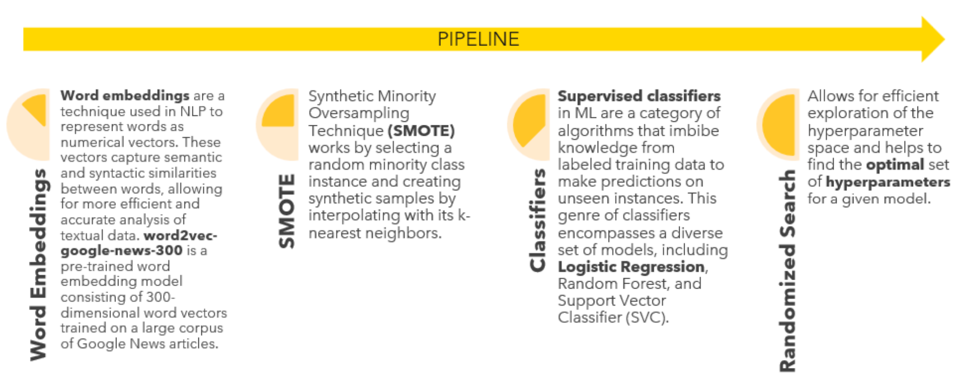


- NER Vector: A one-hot encoding NER vector with labels represented by zeros and ones was created, with a predefined corresponding list of NER labels.
- Length of reviews: Number of words in the text was also incorporated into the model to explore the affect it had on the predictions.

Vectorization:

For converting our text to vectors we have looked at two different approaches here. One is the TFIDF which represents text as a bag of words (i.e., use the frequency of each word to assign weights to each term), and the other is the word embeddings, where the vectors capture semantic relationships between words, and is useful for text classification where meaning is important. On running the model on both, we found that the word embeddings gave us better results (around 3-4% increase in the F-1 score).

We ended up using **word2vec-google-news** pre-trained model. It generates 300-dimensional word vectors for a vocabulary. Using this model was a good starting point for this problem due to its effectiveness in similar contexts. As the model was trained on a large corpus of text from Google News articles covering a broad range of topics, including business and finance the model is likely to have seen a significant number of articles about Axon and its products and may have learned useful embeddings for these terms.



After generating the word embeddings and concatenating them with the additional features, we did SMOTE on our data. SMOTE creates synthetic samples by interpolating with its k- nearest neighbors for the minority classes.

We used cross-validation in our code. The code sets up a k-fold cross-validation with 5 splits and shuffling the data. This technique helps evaluate the model's performance by dividing the dataset into 5 parts and using each part as a validation set once while the rest are used for training. Shuffling ensures randomness, reducing bias and capturing different patterns in the data. This approach provides a reliable estimate of the model's accuracy by assessing it on multiple independent train-test splits.

Then we used Supervised classifiers, we experimented with **Logistic Regression**, Random Forest, and Support Vector Classifier (SVC). While Logistic Regression with L2 Regularization gave us the best results. Further we did randomized search, to find the optimal set of hyperparameters.

HyperParameters used by all the models, along with the probability threshold:

```
Category:  Sales
Model:  LogisticRegression(C=0.75, solver='saga')
Threshold:  0.4
Category:  Cost
Model:  LogisticRegression(C=0.65, multi_class='ovr', solver='newton-cg')
Threshold:  0.55
Category:  Customer Service
Model:  LogisticRegression(C=0.75, solver='sag')
Threshold:  0.4
Category:  Product
Model:  LogisticRegression(C=0.85, multi_class='multinomial', solver='newton-
cg')
Threshold:  0.45
Category:  Training
Model:  LogisticRegression(C=0.75, multi_class='ovr', solver='saga')
Threshold:  0.4
```

Rationale behind these models:

Logistic Regression: Logistic Regression is a popular choice for binary classification problems. When combined with word embeddings, logistic regression can effectively capture the linear relationships between word features and class labels. It is computationally efficient and interpretable, making it a suitable choice for simple text classification tasks.

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. When applied to text classification with word embeddings, It can capture non-linear relationships between word features and class labels, providing good generalization & accuracy.

SVC: SVC is a powerful algorithm used for binary and multiclass classification. When combined with word embeddings, SVC can handle complex text classification tasks by identifying non-linear decision boundaries.

Word embeddings, such as Word2Vec, are vector representations of words in a continuous space. They capture semantic and syntactic relationships between words, allowing the models to leverage this contextual information for classification. Logistic Regression, Random Forest, and SVC are popular choices because they can effectively utilize these word embeddings to learn patterns and make accurate predictions in text classification tasks.

Tools: Jupyter Notebook, Tableau,

Results and Conclusions – Siri Kademani

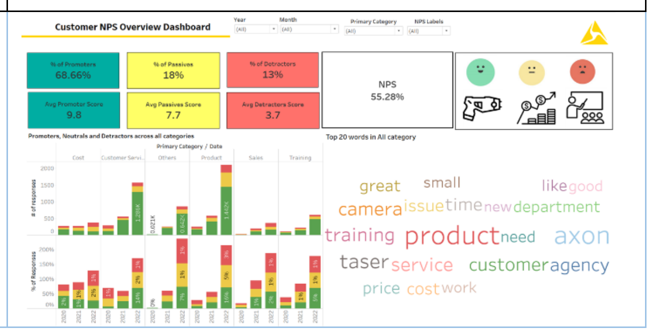
Evaluation metrics used for binary classification model prediction of the 6 categories are:

- **Accuracy**: It represents the proportion of correctly classified instances over the total number of instances in the dataset.

- **Precision**: The probability of making a correct category class classification. It's computed as the number of True Positives divided by the total number of positive calls.
- **Recall**: The percentage of actual ‘Customer Service’ category reviews that were correctly identified as the ‘Customer Service’ category by the model.

Category	Accuracy	Precision	Recall
Cost	93%	86%	86%
Product	88%	79%	86%
Sale	88%	51%	53%
Training	93%	68%	69%
Customer Service	79%	46%	62%

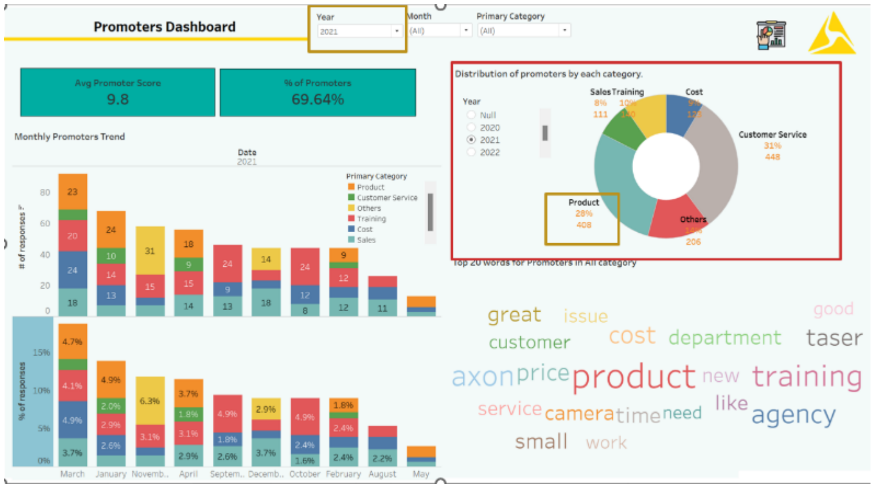
Tableau dashboard to visually represents the effect of model categorization compared to the effect of manual categorization. Data was collected for the years 2020, 2021 and 2022.

Before running the model	**After running the model**
The data was manually classified and was given to us by the client which was mostly prone to human error. This data was around 2K.	The data was fetched on running the model to categorize them into 6 categories (Product, cost, sales, customer service, and others) over 6K uncategorized data which was again merged with the previously classified 2K records.
This was because the manual classification only focused on detractors' comments (with NPS <=6) rather than focusing on all comments.	This merged data represented a much better descent NPS Score of **55%** which included all NPS categories (Promoters, Passives and Detractors).
This lacked in capturing trends and patterns for promoters, detractors, and passives across 6 categories.	This resulted in accurate categorization of NPS labels and helped balance the data, thereby capturing trends across years.
This data gave us the overall NPS Score of only 2%.	
	

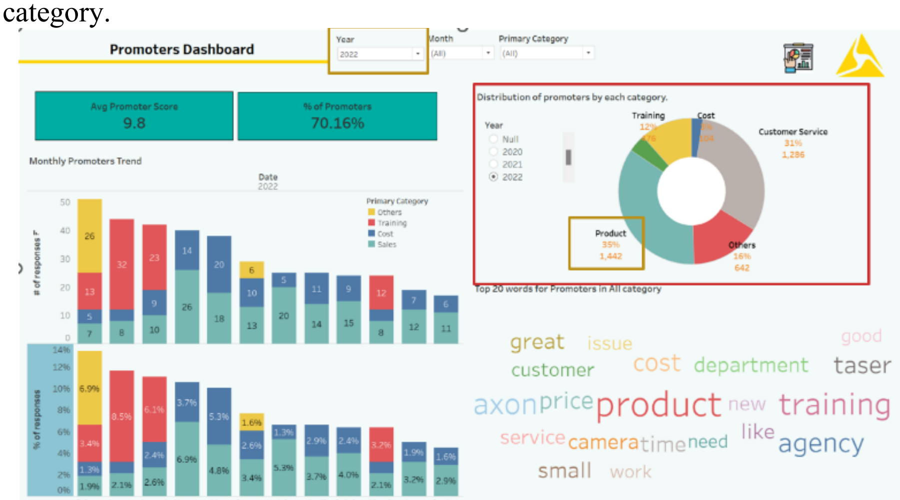
<u>Insights on categories after the model prediction and categorization:</u>

- **‘Product’**

For the year **2021, 28%** of the Promoters belonged to ‘Product’ category.

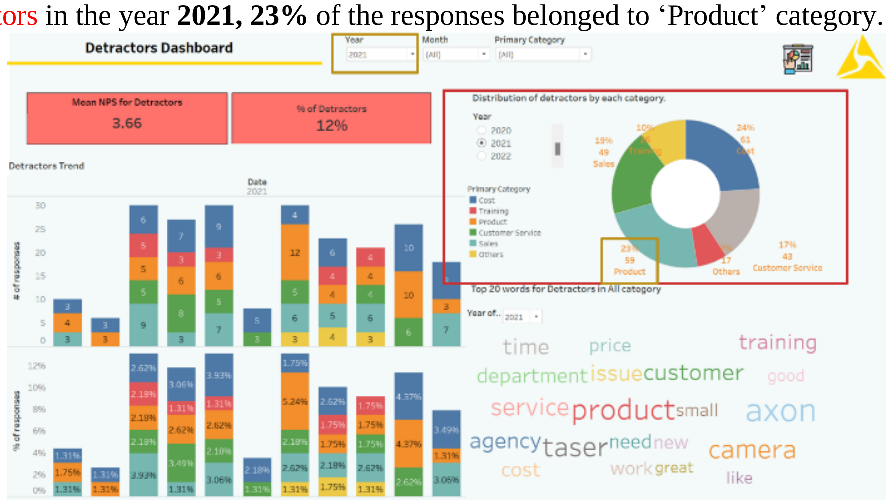


In the Promoters dashboard below, we observe that for the year **2022, 35%** of the promoters belonged to the 'Product' category.

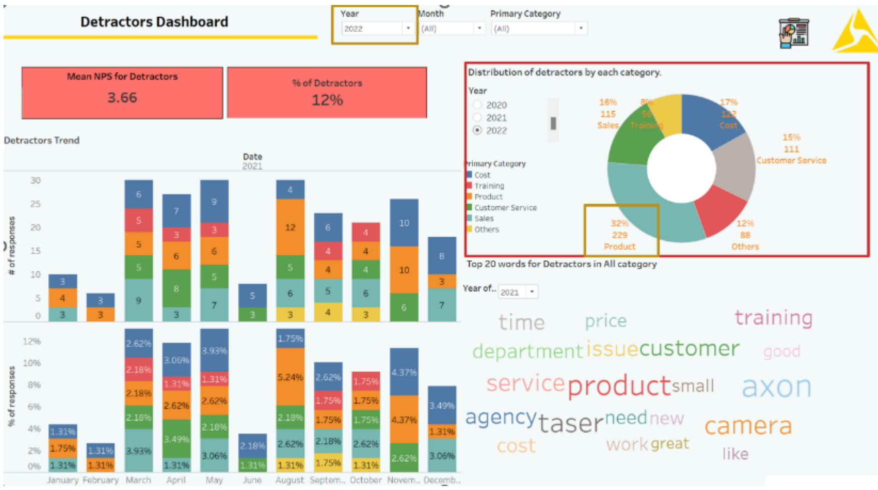


And There was a **7%** increase in the % of Promoters belonging to 'Product' from 2021 to 2022.

But for Detractors in the year **2021, 23%** of the responses belonged to 'Product' category.

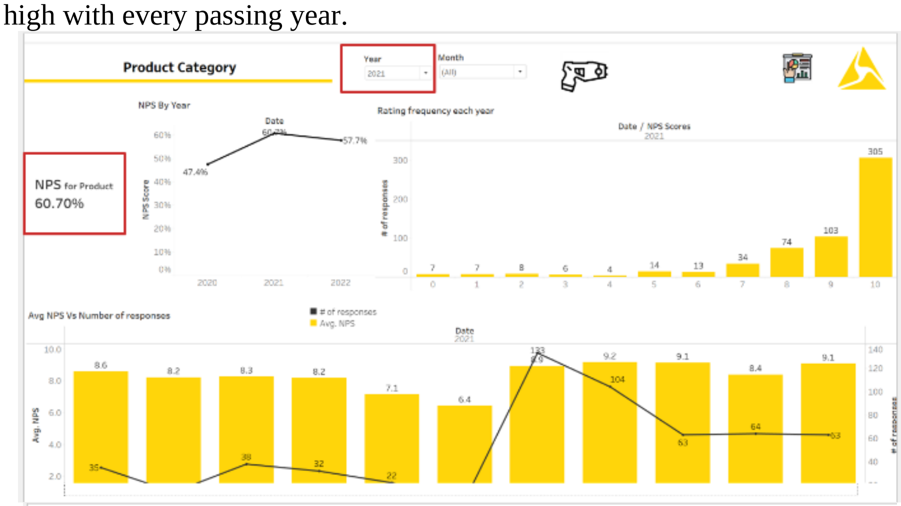


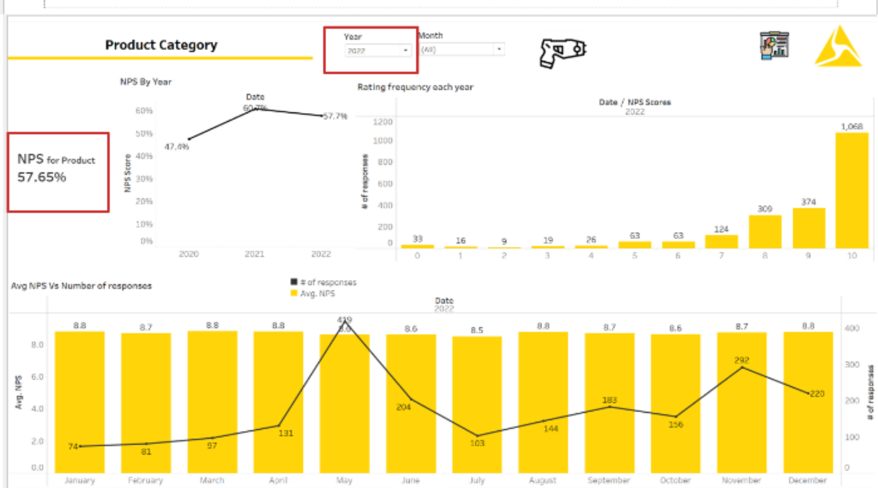
For Detractors, in the year **2022, 32%** of the responses belonged to 'Product' category.



There was a **9%** increase in the % of Detractors belonging to the 'Product' category.

Hence, the NPS for the 'Product' category, for the year **2021 was 61% and for the year 2022, it was 58%**. This accounted for a **decrease of 2%** of the overall NPS for 'Product' from the year 2021 to 2022. This was due to a **significant increase** in the % of Detractors by 9% from the year 2021 to 2022. Hence, it is recommended to focus on comments related to 'Product' and make key improvements in order to keep the NPS high with every passing year.

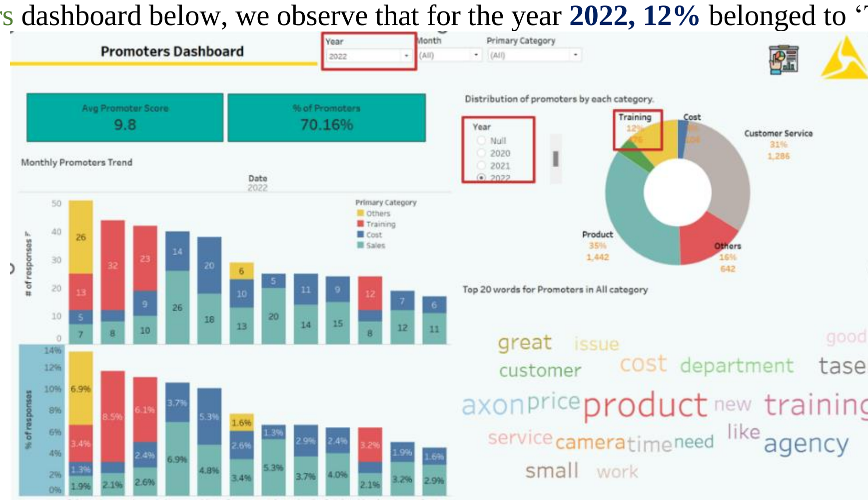




➢ **‘Training’**

For the year **2021, 10%** of the Promoters belonged to ‘Training’ category.

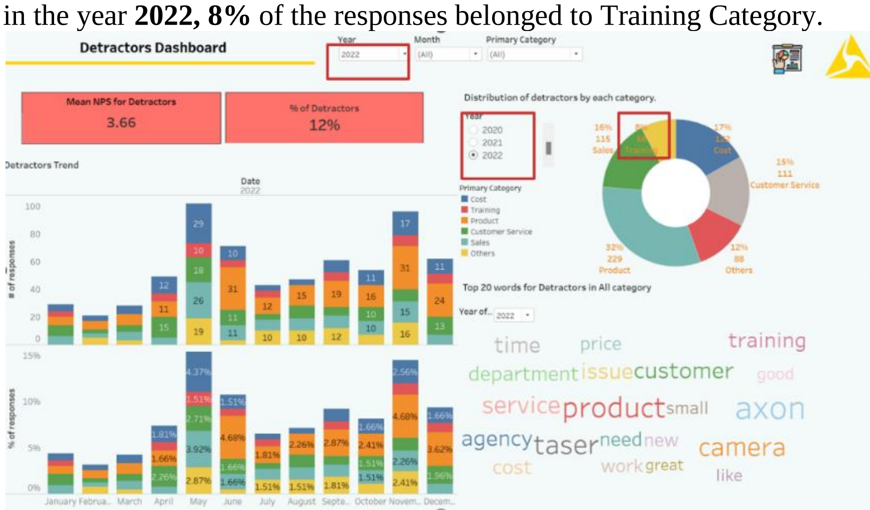
In the Promoters dashboard below, we observe that for the year **2022, 12%** belonged to ‘Training’.



And there was a **2% increase** in the % of Promoters belonging to the ‘Training’ category from 2021 to 2022.

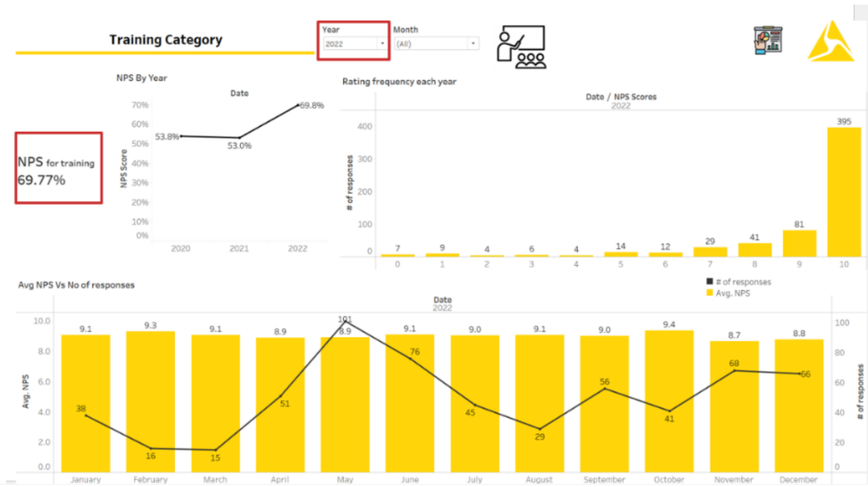
But for Detractors in the year **2021, 10%** of the responses belonged to ‘Training’.

For Detractors, in the year **2022, 8%** of the responses belonged to Training Category.



There was **a 2%** decrease in the % of Detractors belonging to the ‘Training’ category.

Hence, the NPS for ‘Training’ category, for the year **2021 was 53% and for the year 2022, it was 70%**. This accounted for an **increase of 17%** of the overall NPS for ‘Training’ category from the year 2021 to 2022. This was due to an **increase** in the % of Promoters and a decrease in the % of Detractors by 2% from the year 2021 to 2022. Hence, it is recommended to focus on strengths by going through comments related to ‘Training’ and making key improvements to keep the NPS high with every passing year.



- ➢ The final output file will be a csv file that will consist of categorized data. This can be used by the Customer Satisfaction team to identify the strengths and weaknesses of their products and services.
- ➢ The Tableau dashboard can be leveraged by the Customer Experience team to provide key insights on which categories are performing better and which categories need improvement.
- ➢ The suggestions and insights from these dashboards can be used by the higher management to make necessary changes to their products/services.

Model Validation Pre-deployment:

- ➢ **Data collection**: Collect diverse and representative data, which can be achieved by encouraging customers for feedback and simplifying the survey system.
- ➢ **Model Selection**: Select model based on performance, complexity, scalability, interpretability, adaptability, computational efficiency, and robustness. In this case, binary classification had better results than multiclass.
- ➢ **Customer Feedback Validation:** Validate customer feedback against model predictions to obtain insight and identify potential issues like over fitting or addition of features.
- ➢ **Generalization and Scalability**: Test and evaluate model for its ability to generalize well to the diversity of actual customer feedback.
- ➢ **Evaluation metrics**: Identify appropriate metrics (e.g., F1 score, precision, recall, accuracy) to assess the model's accuracy, and scalability for reliable and accurate customer feedback prediction

Model Validation Post-Deployment:

- **Performance Monitoring**: Monitor model performance in real-time to maintain the expected level of performance.
- **Data Monitoring**: Analyze collected customer feedback data to identify potential issues affecting model performance and make necessary adjustments.
- **Continued Analysis**: Compare model performance metrics pre and post deployment to evaluate effectiveness in achieving intended goals.
- **Feedback Analysis**: Gather feedback from users to assess satisfaction and address concerns.

Limitations and Future Scope:

1. **Automation:** The model could be hosted on a server so that when customer feedback is received, it can automatically categorize them based on the categories saved in the database
2. **Dissatisfaction Prediction:** The model could be enhanced to predict dissatisfaction from survey responses for proactive measures to prevent a low NPS score.
3. **Web Application Development:** A web application could be developed to easily access trends by categories, trends by NPS scores, and other relevant information.
4. **Regular Maintenance**: The model should be regularly trained on new customer feedback to ensure its robustness and efficient performance.

References

Liu, X. (2022). Lecture 3 - Text Processing Techniques. CIS 509: Data Mining II, X. Liu, Arizona State University.

Liu, X. (2022). Lecture 5 - Text Classification. CIS 509: Data Mining II, X. Liu, Arizona State University.

Liu, X. (2022). Lecture 6 - Sentiment Analysis. CIS 509: Data Mining II, X. Liu, Arizona State University.

Liu, X. (2022). Lecture 18 - Word2Vec. CIS 509: Data Mining II, X. Liu, Arizona State University.

Appendix – Reproduction of Results – Amit Manocha

Tools Used: Jupyter Notebook, Tableau

All the scripts along with the dataset were put in the folder, which was later archived and sent to the Axon via email.

To use the code in your Python environment, you will need to install the following libraries:

- Pandas: pip install pandas (Version: 1.4.4)
- NumPy: pip install numpy (Version: 1.23.5)
- Seaborn: pip install seaborn (Version: 0.12.2)
- Matplotlib: pip install matplotlib (Version: 3.6.2)
- Plotly: pip install plotly (Version: 5.9.0)
- ipywidgets: pip install ipywidgets (Version: 7.6.5)
- Scikit-learn: pip install scikit-learn (Version: 1.1.2)
- imbalanced-learn (for SMOTE): pip install imbalanced-learn (Version: 0.10.1)
- statsmodels: pip install statsmodels (Version: 0.13.5)
- spaCy: pip install spacy (Version: 3.3.1)

- gensim: pip install gensim (Version: 3.4.0)

Usage Guide

The folder contains 2 jupyter notebook, generate_model.ipynb , generate_output.ipynb,
Takes input from NPS Data for ASU Team.csv

1. **generate_model.ipynb**, needs to run a csv file in Axon format (NPS Data for ASU Team.csv) which consists of classified records. The model trains on the following data file. Tunes hyperparameters and asks for probability threshold for all the six models for each category then saves the model parameters in a dictionary format (model) as a pickle file. The pickle file is saved in the main directory, along with a backup folder is created with the name model_datetime which also contains the model. This file also saves the spacy model (spacy_model_updated_v2), and the word vector file (word2vec-google-news-300) in the main directory.

2. **generate_output.ipynb**, once the model has been generated this file runs and takes the pickle file, the space model, and google vectors as the inputs along with the csv file by the name (NPS Data for ASU Team Output.csv) and generates the output with all the categorized records and saves them as (Categorized_Data_datetime.csv)