Black, Aaron
Chlebowy, Jackson
Loftus, Logan
Sheppard, Maguire

Executive Summary

In a period marked by persistent inflation and shifting consumer behavior, understanding public sentiment toward prices has become a critical challenge for economists, businesses, and policymakers alike. Price volatility directly impacts the budgets of millions of families, influences consumer support and confidence, and shapes policy decisions at the government level. Understanding what economic factors exactly drive changes in consumer sentiment can provide individuals with answers to commonly overlooked behavioral and emotional patterns in response to inflation. Developing a predictive model that integrates economic and behavioral data can help explain how people perceive and react to the constantly changing prices at the store. By comparing these patterns with consumer price indexes and predicting sentiment analysis, our project addresses the gap between economic data and public perception, creating a more responsive, forward-looking understanding of consumer attitudes.

A predictive model designed to anticipate shifts in sentiment surrounding costs can realistically help many different cohorts of people: the government, businesses, and consumers. However, our focus will be on providing these predictive methods and sentiment insights to businesses. As we all know, businesses can utilize numerous forecasting techniques to price their products and gain a competitive advantage in the market. With the data we will collect, our analysis will take some of this information and translate it into insights that help with business forecasting.

To accurately build a predictive model capturing public sentiment regarding price fluctuations, several best practices must be followed. These practices span data collection, sentiment definition, model selection, and interpretation ensuring that the final output is both reliable and actionable for businesses and consumers.

The first step is to determine how the sentiment target is defined in the model. We could group the sentiment into discrete categories such as good, neutral, and bad. This approach marks the problem as a classification task, allowing the model to label regions or time periods based on sentiment status. Or we could express sentiment as a continuous score, this labels the problem as a regression task, enabling us to quantify sentiment intensity over time or across regions. To effectively practice sentiment prediction requires collecting the correct and relevant data, combining both structured and unstructured data. We explain the data and how we plan to collect it later in this document. The best practice to apply sentiment analysis is to use one of the most relevant tools, such as NLTK's VADER sentiment lexicon. We can use VADER to score each post or article's sentiment and then apply the appropriate ML techniques based on the specific sentiment format Logistic Regression and Decision Tree-based method for classification.

The data we plan to analyze contains a time-series component, indicating that temporal dependencies play in this prediction task. For our analysis, the granularity of our data is by year, month, and state, making our model helpful in predicting future consumer sentiment.

Structured data that is crucial for predicting local sentiment from changes in prices are Consumer Price Index (CPI) along with monthly and regional based fluctuations. Other forms of data that can be used to predict local consumer sentiment are unemployment percentages, crime rate, and other societal factors. This data can be accessed through the Bureau of Labor Statistics (BLS), which is a publicly available data source. We can easily export the BLS data into a structured CSV format for data analysis, therefore there should be no feasibility issues with this data collection.

The unstructured data comes in the form of local sentiment of the communities affected by the price changes. The best source for this data that we will use is either Reddit or Twitter because we can use a natural language processing sentiment analysis to figure out the attitudes being expressed. This data is easily accessible since Reddit is a popular social media platform. However, the collection process will be time consuming since we will need to use API scraping for a sufficient amount of responses. Additionally, we are assuming that every Reddit post we collect is legitimate and honest.

The results from this project were that the most influential predictor was disposable income in our overall random forest model, with a model accuracy of 0.965. The state level random forest had an accuracy of 0.63 with the most important variable being produce CPI. The logistic regression accurately predicted 73 percent of all observations. The insights we gained were that CPI is not the only useful feature in predicting consumer sentiment. We found it interesting that the state level sentiment had different feature importances than the overall Midwest. This can be attributed to states having different feelings towards the current economic standing.

Aside from business applications, our model and public sentiment information could create awareness of how socioeconomic/political factors affect consumers and inform policymakers on how to make meaningful changes in making prices changes and other living costs more affordable for the everyday consumer. The government could also apply our model and sentiment information to address issues that directly affect consumers' sentiment to inspire confidence and limit public panic.

In conclusion, the goal of this project is to bridge the gap between economic indicators and consumer perception to predict public sentiment using a combination of economic, social, and behavioral data. By integrating structured data like the consumer price index and unstructured social media sentiment analysis, our model will provide a more holistic look on how consumers experience changes in affordability. The insights generated from our model can help inform business strategy and decision-making, ensuring companies can adapt more effectively to changing consumer needs and economic conditions.