

Neil Bajaj

Graduating senior from University of California, San Diego (UCSD) - June 2024

Compilation of UCSD Projects

- Company stock and Layoff Performance
- Credit Card Application
- Predicting Amazon Review Ratings
- Jelly Bean Themed Fortune Website
- Twitter Sentiment Analysis
- Deloitte-UCSD Case Study

Email: neilbajaj1102@gmail.com, Phone: +1 (858) 386-2046

Company Stock Performance and Layoff Decisions

Goal:

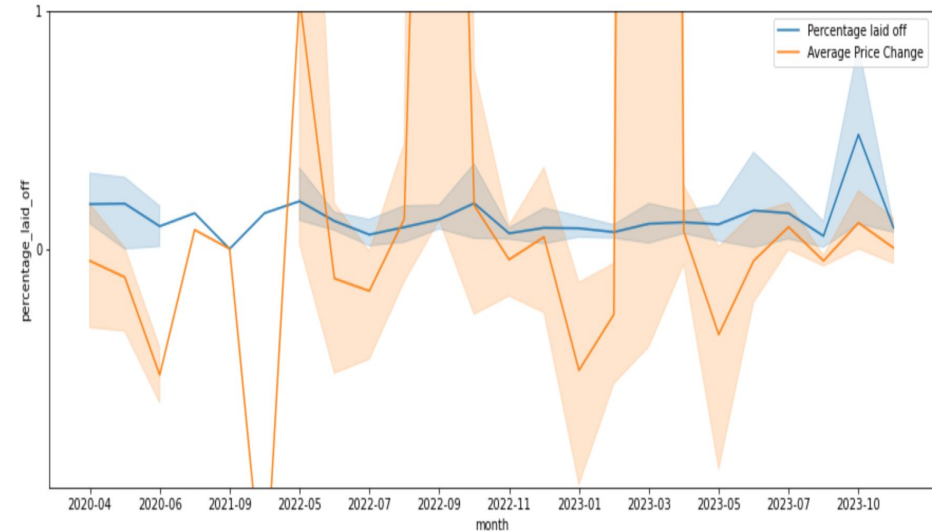
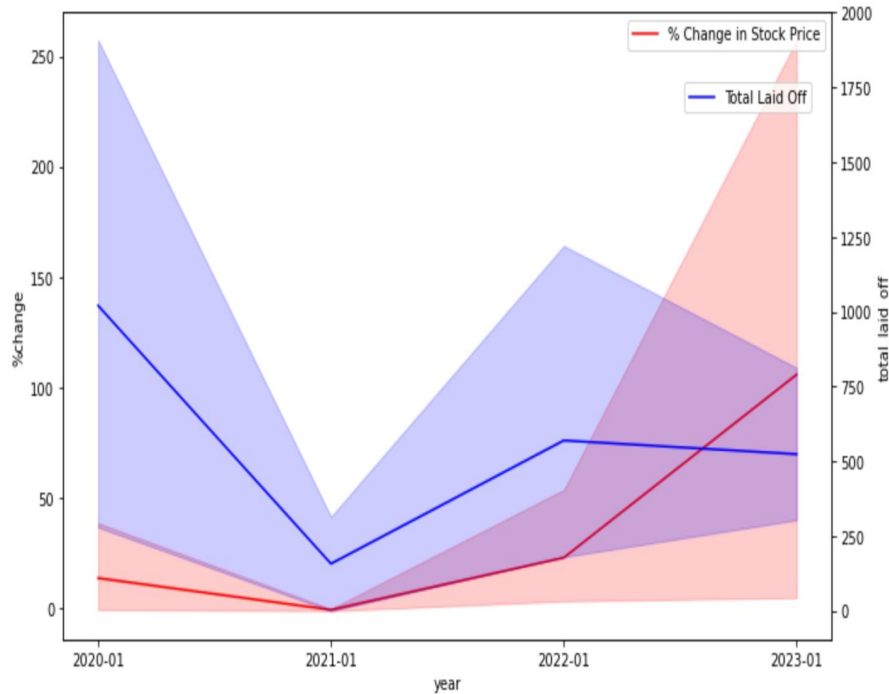
To determine if Stock Prices of technology Companies in United States have a correlation with their layoff decisions from the year 2020 to 2023.

- Merged Datasets to create line plots to determine if trends in stock performance directly led to layoff decisions within that company.
- Analyzed Pearson correlation and P- values to determine the result. Determined if the correlation is significant if the P-value is below a certain threshold (0.05).

Results:

- Found no direct correlation between stock market price changes and layoffs in the tech sector, suggesting complex interactions between financial performance and employment strategies.
- Potential confounders include variations in employee performance, shifts in project interests, and overall company performance.

Results/ Visualizations - Stock/Layoff



- Layoffs line fluctuates, indicating dynamic workforce trends; stock price line reflects companies' financial performance.
- Correlation between the two lines offers nuanced insights into the relationship between layoffs and stock price shifts in tech industries.

Credit Card Application Approval

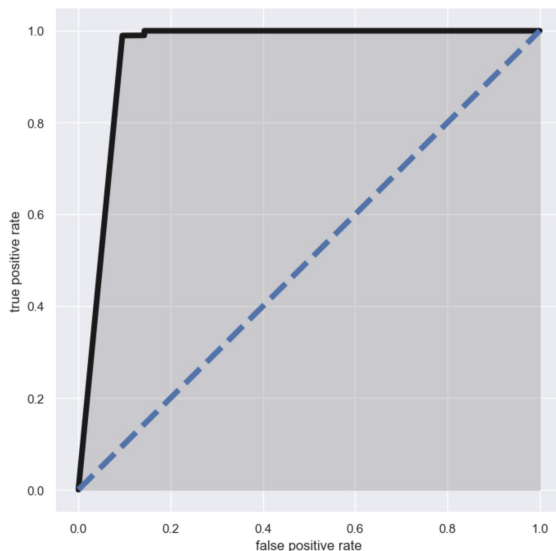
- Implemented Supervised Machine Learning Techniques to build a predictive classifier that determines whether a credit card application would be approved or rejected based on the applicant's history.
- Tested and Deployed models like:
 - Logistic Classifiers
 - Gradient Boosting
 - Ada Boosting
 - K-NN Classifier
 - Random Forests Classifier
 - K-fold cross Validation
 - repeated K-fold Cross Validation



Credit Card Application Results

Visualizations for the results of using a Logistic Regression Classifier to determine whether the credit card applications gets approved or declined.

It gave a high true positive rate and an accuracy score of 99%.



Grid search

Below are the screenshots of the Gridsearch for our logistic classifier over the hyperparameters

```
search_space = {'classifier__C': np.logspace(-4, 4, 9),
                'classifier__penalty': ['l2', 'l1', 'elasticnet'],
                'classifier__solver': ['saga', 'lbfgs', 'newton-cg', 'sag', 'newton-cholesky', 'liblinear']}
```

```
# Create a grid search object to find the best model
best_model = GridSearchCV(pipe, search_space, cv=5, verbose=3, n_jobs=-1)
```

```
In [180]: #FITTING OUR BEST MODEL
          #FINDING THE BEST ACCURACY
          #why are we splitting again here? shouldnt we fit on res anyways

          # Fit grid search
          best_model.fit(X_res, y_res)
```

The gridsearch resulted in a best model with its parameters given below

```
best_model.best_params_, best_model.best_score_ #maybe too good?
```

```
Out[182]: ({'classifier__C': 10000.0,
            'classifier__penalty': 'l2',
            'classifier__solver': 'lbfgs'},
           1.0)
```

```
#LOGISTIC REGRESSION BEST MODEL AND ACCURACY SCORE
clf = LogisticRegression(solver='lbfgs', random_state=42, C=10000, penalty='l2').fit(X_res, y_res)
score = clf.score(X_test_df, y_test)
print(score)
```

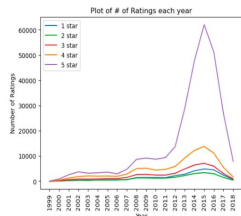
0.9972144846796658

Predicting Amazon Review Ratings

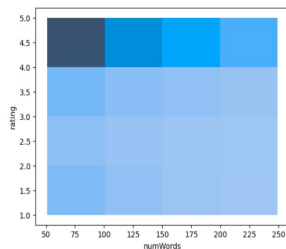
- Developed a web mining and recommender system using Amazon's Video Game Reviews and Ratings dataset, predicting ratings with a low MSE of 1.07.
- Executed data cleaning, visualization, and analysis in Jupyter Notebooks to build k-nearest neighbors, various latent factor models, linear regression, and SVD models using NumPy, Sklearn, matplotlib, pandas, and TensorFlow.



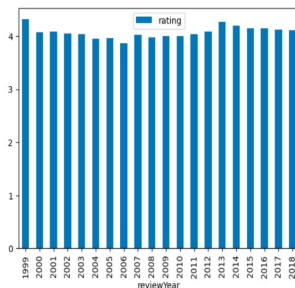
Interesting Findings - Amazon Reviews



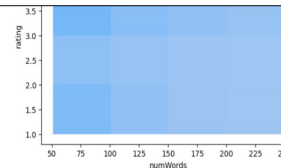
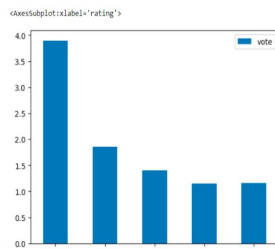
From the above plot, we can see that the number of reviews has increased over time, and also that 5 star ratings are the most popular ratings of all ratings.



Above is a heatmap of rating based on number of words. It is clear that 5 star ratings is still the most common even as the number of words increases.



Average rating of all reviews through the years. As shown, it does not fluctuate greatly and remains around a rating slightly above 4.

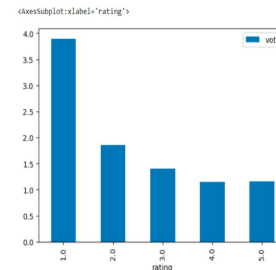


Above is a heatmap of rating based on number of words. It is clear that 5 star ratings is still the most common even as the number of words increases.

Interesting findings -

- there are more ratings with value five than ratings with the other four values combined.
- The review with the maximum length was 6000 words but 75th percentile of word lengths was length 350.
- Ratings with values 1 and 5 have lower words per rating than ratings with values 3, 4 and 5.
- Ratings in the year 2018 had the lowest number of words per rating on average
- Ratings with value 1 received more votes than any other ratings.

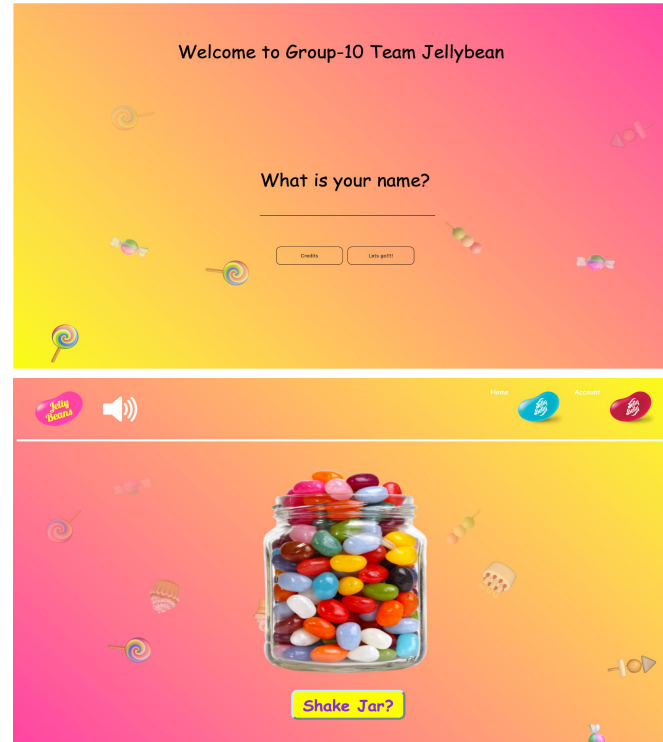
above 4.



Average number of helpful votes for each rating. It is interesting to see that ratings with 1 have the most amount of votes on average.

Jelly Bean Themed Fortune Telling Website

- Designed and deployed a Jelly Bean themed fortune telling website built using HTML, CSS and Javascript which uses Generative AI through Open AI API to generate and store daily fortunes for every user.
- Used Agile workflows (sprint planning, sprints, sprint-retrospective, user stories, backlog, CI/CD) within our team to increase efficiency and reduce the development time by more than 50%.



The following pictures represent the UI of the website and showcases how it works.

Integrated a CI/CD pipeline, consisting of a Node JS Tester, Prettier, Super Linter and JS Docs documentation, to improve and automate the software delivery process as well as to identify possible bugs in our web.



[Home:](#)  [Back:](#)  [Account:](#) 



Your daily fortune is:

Silver represents intuition. Today, trust your instincts and make decisions with confidence.



Hi nb, Welcome to your account page! Here you will find all your past fortunes

You shook the jar 1 times!

You have picked 4 jellybeans!

With a touch of blue, your day will be filled with serenity and harmony. Enjoy the peaceful moments. Green brings good fortune! Today, luck will be on your side, bringing you unexpected joy and surprises. Orange ignites enthusiasm and creativity. Today, approach your tasks with passion and innovative thinking. Embrace the festive color of margarita! Today, let your spirit soar and embrace the lively energy around you.

Index	Jellybean	Fortune
1	blue	With a touch of blue, your day will be filled with serenity and harmony. Enjoy the peaceful moments.
2	green	Green brings good fortune! Today, luck will be on your side, bringing you unexpected joy and surprises.
3	orange	Orange ignites enthusiasm and creativity. Today, approach your tasks with passion and innovative thinking.
4	margarita	Embrace the festive color of margarita! Today, let your spirit soar and embrace the lively energy around you.

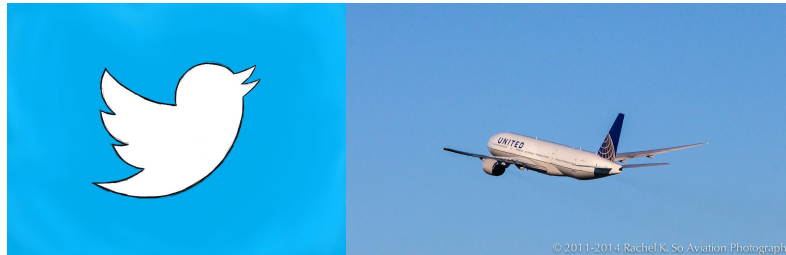


Twitter Sentiments Analysis

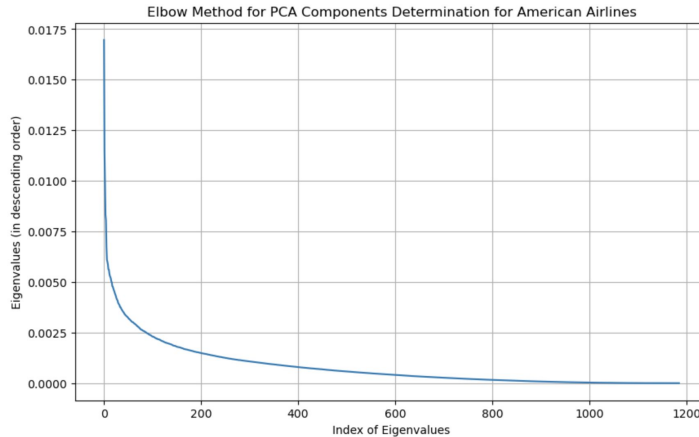
Goal:

Implemented Sentiment Analysis to analyze the sentiment of customer satisfaction tweets towards Airlines operating in the United States of America.

- Utilized NLP and clustering techniques to group tweets targeting airlines (United, American, Southwest, Virgin America, US Airways, Delta) into distinct categories using K-Means and Gaussian Mixture models.
- Employed Latent Dirichlet Allocation for topic modeling to analyze reasons behind negative tweets aimed at these airlines.
- Used a silhouette score metric to evaluate the quality of our clustering and results.



Sentiment Analysis Results



The following visualization is for American Airlines to reduce the dimensionality while converting tweets to vector. Similarly was performed for all airlines.

- The study applied natural language processing and machine learning to categorize US airline customer feedback from tweets, identifying key areas for service improvement.
- A sentiment analysis model was developed to detect sentiments in new tweets, despite challenges like low silhouette scores indicating mixed sentiments in clustering algorithms.
- This research highlights the broad applicability of social media feedback analysis for gaining strategic insights and improving service quality across various industries.

Deloitte Mentorship Program with UCSD

Case Study:
Department of Labor
proposal

Goal: What programs can be implemented in the regions with the most struggling industries that could help them get back to pre-covid statistics?

- Used Data Visualization techniques to find out the top Industries that were impacted the most by Covid-19.
- Split the USA map into regions and used heatmaps to plot the percentage change in labor in each of the regions.
- Suggested programs to implement to deal with the rapid decline in the industries around the world.

Link to Case Study project:

<https://docs.google.com/presentation/d/1Y-oBU1wU5VMukynbR5OiGwvdjNlosyxa5MqLSEZhT50/edit?usp=sharing>