

Connect & explore how I can be of an asset to you  
as I have been for others

**# OpenToOpportunities**

## I Implement **Data** and **AI** solutions for Your Use Cases

- ★ Implemented 14+ use cases
- ★ Consulted 4+ Startups
- ★ Regularly invited as panelist
- ★ Speaker at Tech Talks




**Data Analyst --> ML Modeling --> AI Engineer**  
Design Thinking + Rapid Prototyping + Productionize + Story Teller

Data Scientist | Time Series Analytics, Recommendation Systems, Python, SQL, Power BI | I Help  
companies leverage AI, ML & Big Data to Identify Anomalies, Mitigate Risk & Personalize offerings



Get Food Recipe [\[LINK\]](#) (LLM application) [\[code\]](#)

 **Get Food Recipe**

Get Food Recipe based on Grocery, Time Required, Cuisines and Equipment available (all inputs are optional)

Enter Google Generative AI API KEY (Required)

Click for API KEY (select create api key in new project)

Upload an image of ingredients/ utensils/ food...

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Browse files

Particular Food in Mind (Dal Tadka, cake)

Grocery (onion, garam masala)

Cooking Time (1 hr, 30 mins)

Cuisine (Italian, South-Indian)

Equipment used (frying pan, spatula)


Meal (breakfast, brunch)

Preference (vegan, no meat)


Allergies

Additional information/requests

Recipe

Made with  by Amogh Mahadev kokari © 2024 || [linkedin](#) || [Portfolio](#) || [Github](#)

Get Fitness Routine [\[LINK\]](#) (LLM application) [\[code\]](#)

 **Get My Exercise Routine**

Get Weekly Exercise routine based on Goal and target, Time Required, Cuisines and Equipment available (all inputs are optional)

Enter Google Generative AI API KEY (Required)

Click for API KEY (select create api key in new project)

Upload an image of activity/ equipment/ place...

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Browse files

What is your main health goal?

'Lose 5 kg Weight', 'Feel Stronger', 'become lean', 'Have More Energy', 'Feel Better Overall', 'Other'

How intense do you want your physical activities to be? (1-10)

110

How much time can you dedicate each day? eg. 30 minutes

How many days a week can you commit

17

What kinds of physical activities do you enjoy?

Choose an option

Do you have any injuries or health concerns? If yes, please specify

Preference

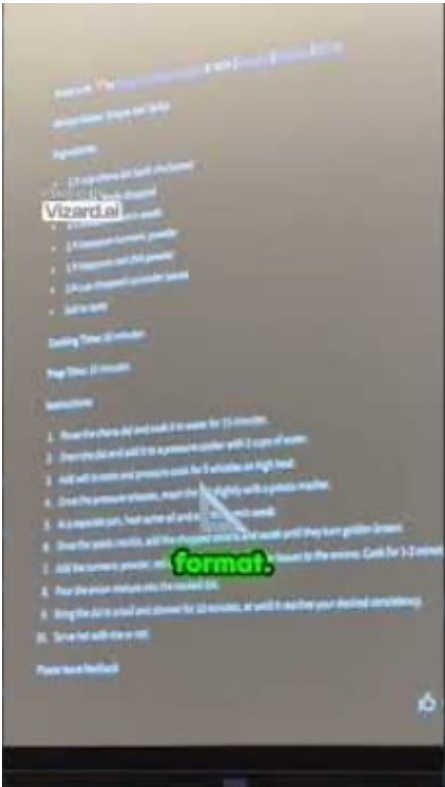
indoor

Any equipment like dumble, weights or ball you want to use

Do you play or like to start any sports or outdoor activity like tennis, Biking

Any other information or preferences like 30 days target, you want to share?

Get Physical Activity Plan



Talk at Microsoft office NYC  
during  
AI Camp Meetup  
[link](#)

## Motivation

- The motivation behind this is to empower the creation of high value content on streaming service like YouTube.
- The underlying fundamental is to build a scalable decentralized framework for ingesting data from different sources, training ML algorithms and predicting user content
- Scope is extended to show the effectiveness of Spark framework on large datasets

## Technology

- PYSPARK
- PYTHON
- FLASK
- PANDAS
- HTML/CSS
- JOBLIB
- LIGHTGBM

# Big Data Text Analysis - YouTube Channel Video Classification

[\[Project Slides\]](#)

[\[DEMO\]](#)

## Workflow

### Enter Youtube Channel URL

Channel URL

### Enter Youtube Channel URL

Channel URL

### Enter Youtube Channel Title and Description

Channel Title   
Channel Description

Classifier Name	Value
GaussianNB	0
LGBMClassifier	1
XGBClassifier	0
AdaBoostClassifier	1

### Word Clouds on all the video data present in the Channel



### ML Algorithms and Analysis Results

Classifier Name	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
GaussianNB	75.87	75.89	75.87	75.86
LGBMClassifier	80.95	80.96	80.95	80.95
XGBClassifier	80.51	80.51	80.51	80.51
AdaBoostClassifier	76.26	76.27	76.26	76.26

### Spark Analysis Results

Classifier Name	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
LogisticRegression	82.36	84.25	93.68	81.15
GBMClassifier	84.17	86.42	93.71	83.32
NaiveBayes	80.88	92.23	81.84	81.77

### Run time Analysis

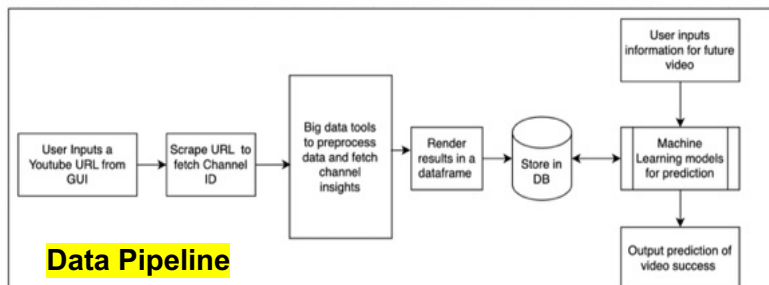
Normal (seconds)	Spark (seconds)
376.81	192.06

### Data shape and class size

Rows x Cols	Rows (R)	Cols (C)
(9055, 3)	4540	4515

Click below to predict how many likes your video might get: [Click Me](#)

## Data Pipeline



## Conclusion

- The end-to-end analytics as a service based on decentralized microservices can be used in creating a robust real time prediction that can help content creators attract crowd on YouTube
- Although the limitation to the current development is on untuned algorithm but they tend to perform well in binary classification (0,1), auto tuning can be applied to make them better
- we believe the proposed big data architecture will be widely extensible and expandable that can be produced as industry grade service for everyone

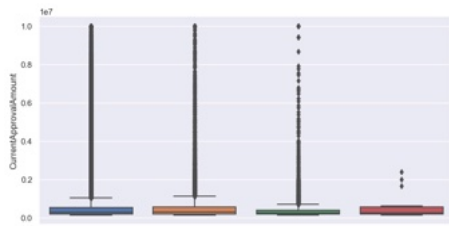
## Current & Future Work

- The current scope consists of analyzing data from one channel from YouTube to train on ML algorithms and Spark based to create predictions from user and display result on Frontend using Flask
- The current scope of deriving insights from one channel can be expanded to multiple channel input
- The preprocessing and ML modeling can be parallelized by running multiple instances
- Decentralization and guarding it from short term failure can be resolved using AWS ecosystem the entire project can be cloud based to handle high volume of requests and process them in real time paving the way of real time streaming analytics and prediction

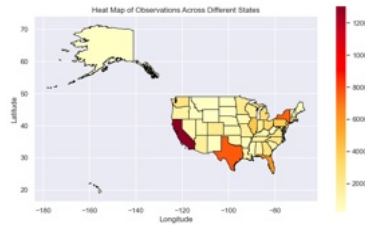
# Fraud Detection - Paycheck Protection Program

## Exploratory Analysis

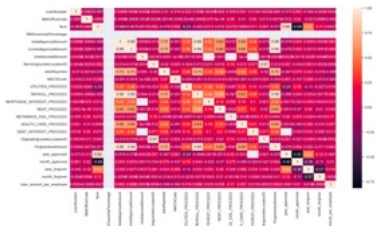
- Investigated data to understand distribution, anomaly with heat map and charts
  - Others were specific to data exploration with Python
- Some examples of this effort can be seen below.



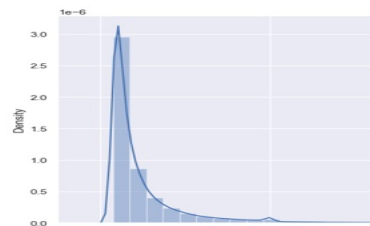
Approval Amount Outlier Analysis



Heatmap of number of loans approved across all the states



Correlation heat-matrix to identify key dependencies



Loan Approval Amounts Density Plot

## Motivation

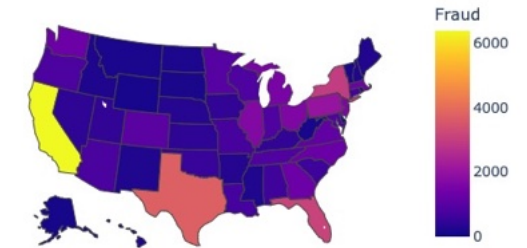
- Explore loan data from the Paycheck Protection Program (PPP), which provided relief to small and medium-sized businesses during the COVID-19 pandemic and identify probable fraudulent loans.
- The primary objective is to reduce frauds in the future by applying anomaly detection methods to identify outliers and building machine learning models to potentially detect possible frauds.

## Anomaly Detection

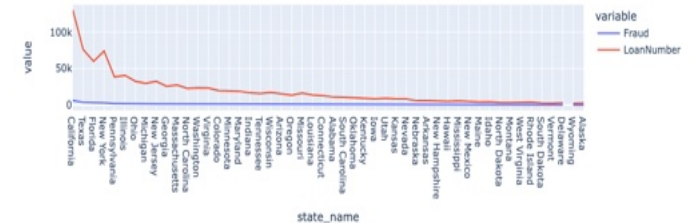
- Data is Standardization using standard scalar
- Findings Higher the anomaly score, higher the probability of loan being fraud
- Used Isolation forest to calculate anomaly scores
- Calculated Anomaly score for features:
  - Initial Approval Amount
  - Payroll proceed
  - Jobs Reported
- Identified loan being fraud biases of anomaly score top 95% = anomaly (i.e. fraud)

## Technology

- Python was used for data pre-processing and cleaning
- Python libraries matplotlib, plotly, seaborn utilized for exploratory analysis and dashboarding
- Python to identify anomalies using pandas, matplotlib, IsolationForest in Sklearn



State-wise probable frauds in USA



State-wise Loans vs Frauds

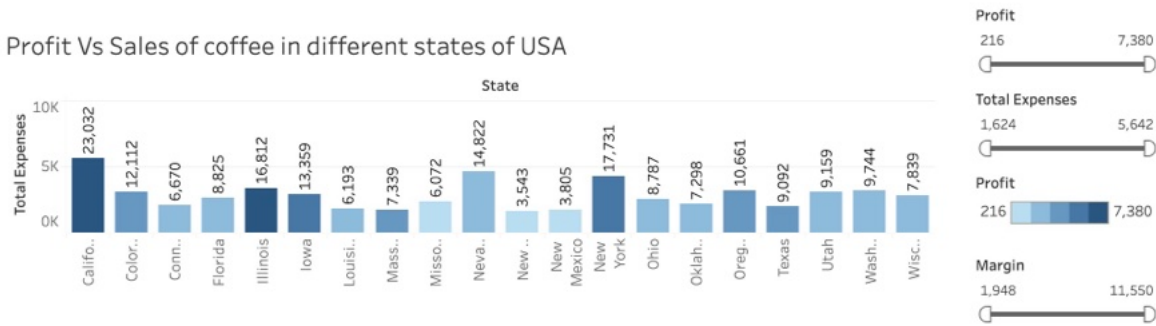
## Current & Future Work

- Government agencies and policymakers may use the data to evaluate the effectiveness of the program and make necessary adjustments to future relief efforts
- Researchers may use data to study the economic impact of the pandemic on small businesses and to identify patterns and trends in the distribution of PPP loans
- Incorporate possible frauds with portal like Datamerch, Experian, lapps to improve due diligence for debt-based Venture Capitalists

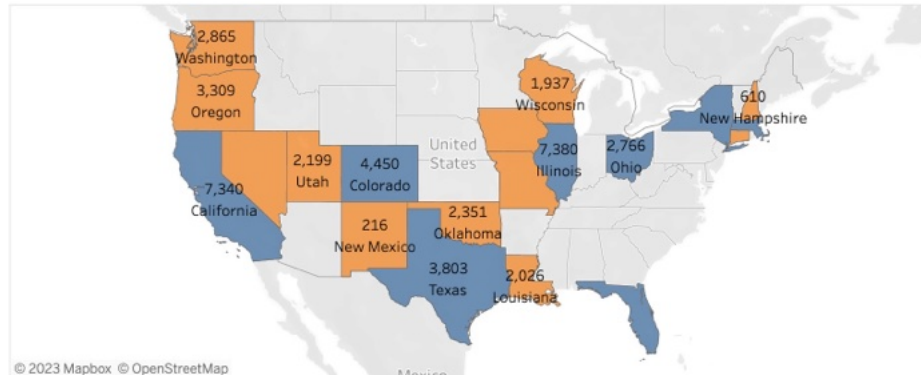


## Tableau Viz on Profit VS Sales of Coffee in Different States of USA [\[LINK\]](#)

Profit Vs Sales of coffee in different states of USA

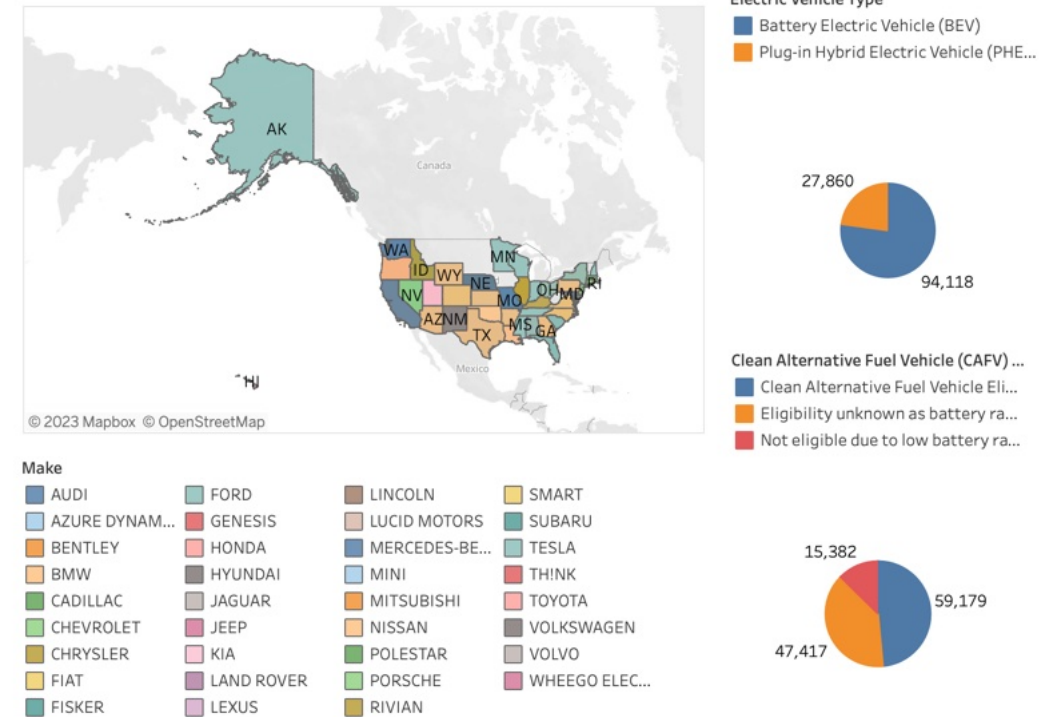


Map



## Electric Vehicle Analysis [\[LINK\]](#)

Electric Vehicle Population Data [Washington US]



## Goal Based Student Diet Personalization

(Interviewed students using **Design Thinking** tools (Empathy), analyzed data to understand user needs to develop product to save busy student's deteriorating health)

[\[Prototype\]](#)

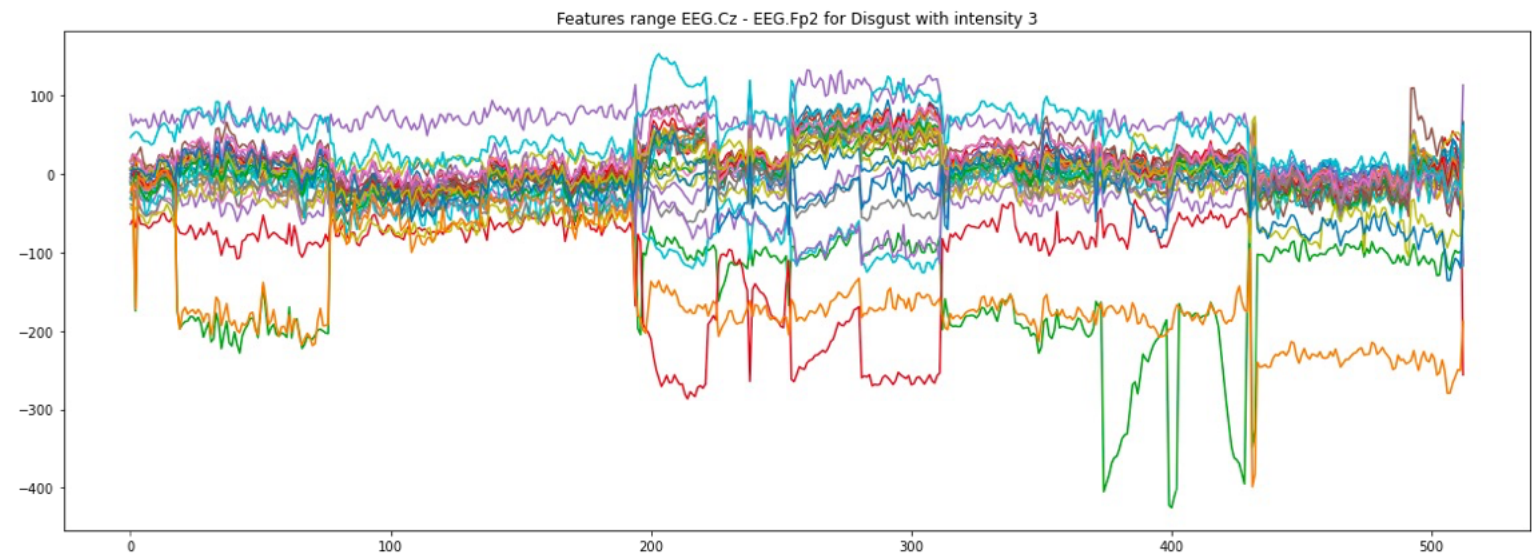
[\[Project Slides\]](#)

[\[Explainer Video\]](#)

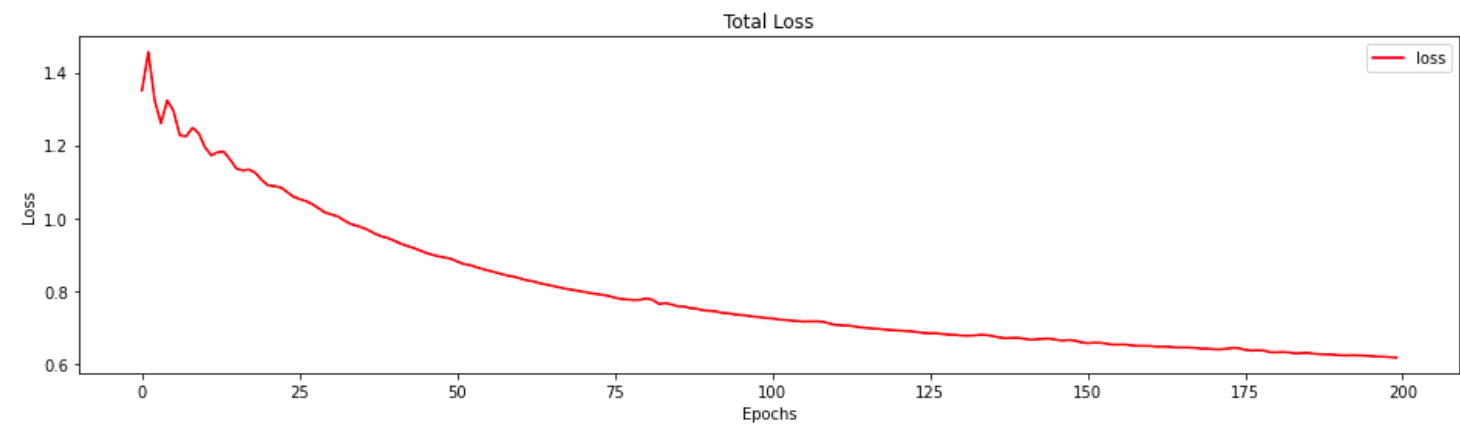


# EEG Emotion Detection

## Deep Learning Hard Parameter Sharing



	precision	recall	f1-score	support
0	0.80	0.77	0.78	13005
1	0.81	0.81	0.81	40448
2	0.81	0.80	0.80	16179
3	0.79	0.81	0.80	54272
4	0.76	0.79	0.77	22733
5	0.80	0.80	0.80	33280
6	0.78	0.74	0.76	17408



accuracy			0.80	197325
macro avg	0.79	0.79	0.79	197325
weighted avg	0.80	0.80	0.80	197325

# Testimonials

## Chinmay Rathod (Co-Founder, Lead Data Scientist)

- I had the pleasure of working with Amogh on a beta testing project that involved deploying code on AWS. His strong knowledge of cloud computing, software development, and data science was incredibly helpful for the project .....

## Balaji Rao (PhD, Blockchain Researcher)

- I worked closely with Amogh for over a year as a research assistant at the Stevens Institute of Technology. Amogh is a highly skilled individual with expertise in software development, data engineering, and data science. He built the Human-Computer Interaction Lab's official website, collaborating with a team of 5 researchers and developers to create an interactive user experience .....

## References available on request

**Michael Washington**  
Director of Finance

**Dr Thomas Lechlar**  
Project Management Professional

**Dr Edward Stohr**  
BIA Program Director

**RJ Lehman**  
Founder Bullwhip

**Dr Joseph Morabito**  
Business Intelligence Professional



# Thank You

[LINKEDIN.COM/IN/AMOGHKOKARI/](https://www.linkedin.com/in/amoghkokari/)

[GITHUB.COM/AMOGHKOKARI](https://github.com/amoghkokari)

[GETFOODRECIPE.STREAMLIT.APP](https://getfoodrecipe.streamlit.app)

