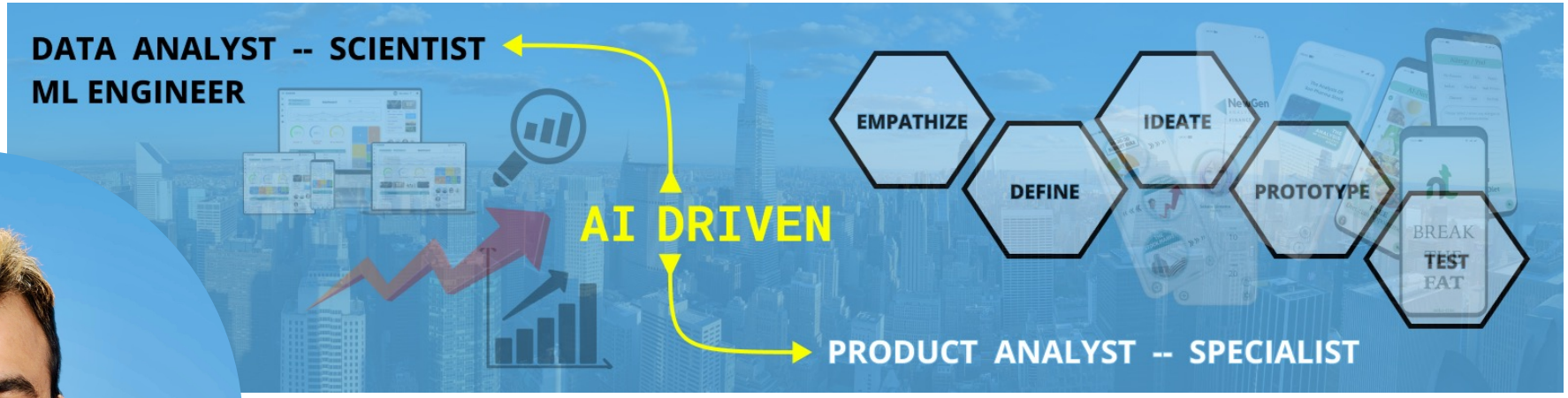


Deriving Data Backed Insights for Informed Decision-Making

Galvanizing User-Centered Growth

aws certified Cloud Practitioner FOUNDATIONAL

Google coursera
Google Project Management Certificate
Certificate of Completion

git

databricks

pandas

Figma

SQL

C3.ai

Power BI

AMOGH MAHADEV KOKARI -- LINKEDIN.COM/IN/AMOGHKOKARI

About Amogh

THE STORYTELLER

Hi there, I am Amogh, a versatile **Data** professional specialized in leveraging **Advanced Statistical Techniques**, **Programming Skills**, **Generative AI** capabilities and domain knowledge to extract actionable insights from complex datasets.

With a Master's degree in **Information Systems** and a proven track record of success in Finance and various other industries, I excel at solving business challenges through **Data-Driven Decision-Making**.

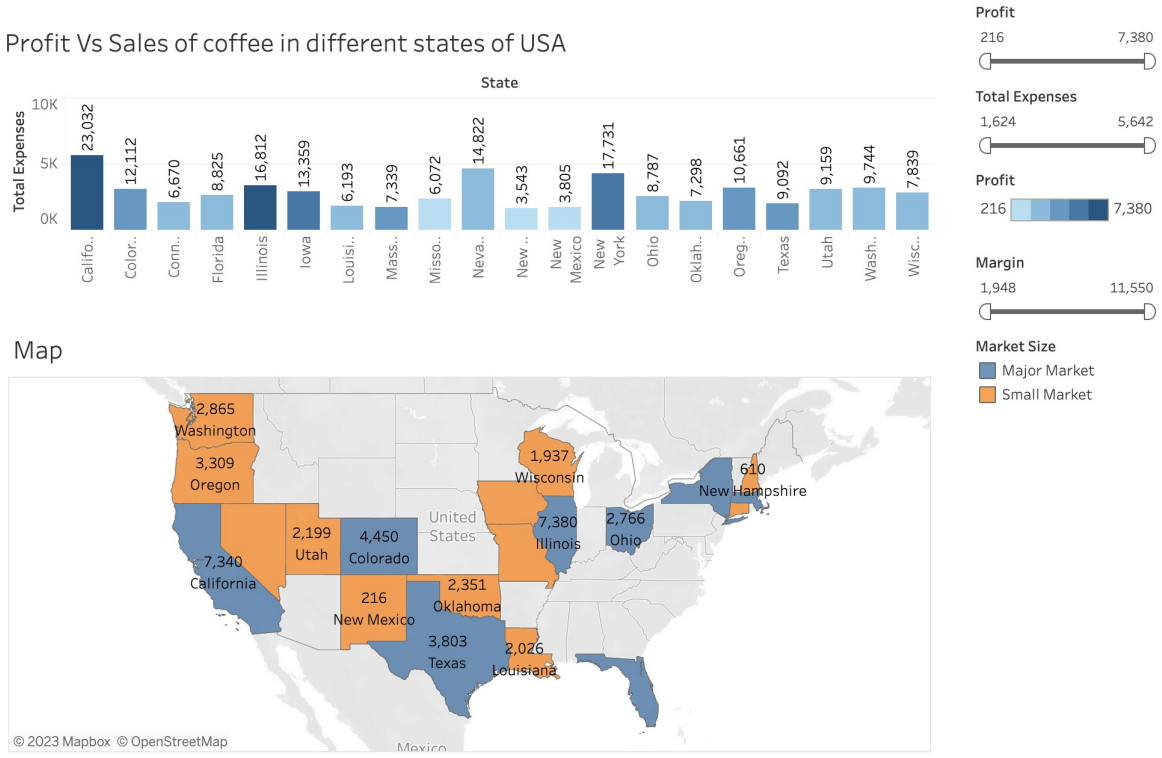
Going through the **Design Thinking** cycle from Ideation to Deployment, I find the solution for **User Centered Problems**.

From developing **Time Series Forecasting** models and **Recommendation Systems** to streamlining processes and automating data collection, I have consistently delivered transformative results.

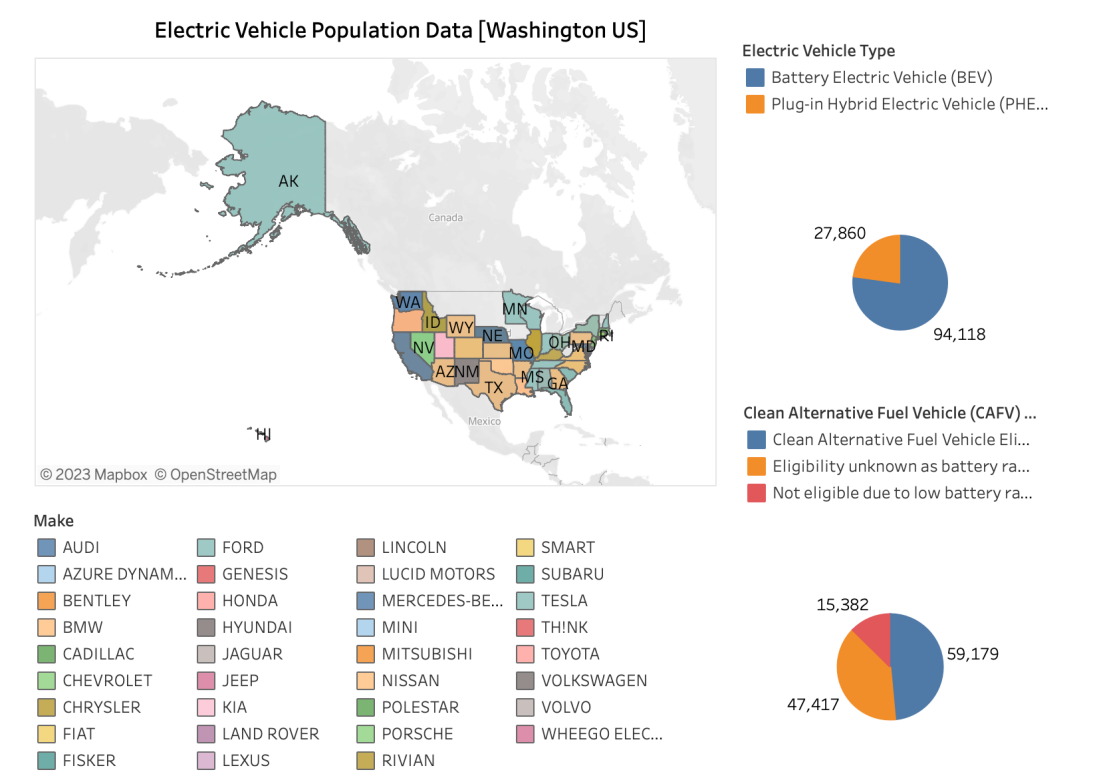
Proficient in **EXCEL, POWER BI, SQL**, **PYTHON** and various analytics tools like Jupyter & Databricks, I possess a strong foundation in **Data Analytics**, Machine Learning, visualization, hypothesis testing, A/B testing, and statistical modeling.

Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

Tableau Viz on Profit VS Sales of Coffee in Different States of USA [LINK]



Electric Vehicle Analysis [LINK]



Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

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.

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 slkearn

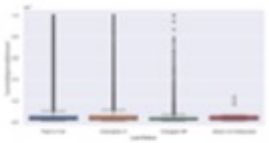
Fraud Detection - Paycheck Protection Program

Exploratory Analysis

Visualizations for this project are intended to aid the user get an essential understanding of the context of the subject and get a feel of what the dataset is trying to convey on the high level.

Supplementary intentions with the visualizations are to be informative for the user to know where to start looking for anomalies first. Ultimately, we want the the visualizations to aid the process of identifying anomalous phenomena by guiding the user's attention to some extreme unruly observations as a starting place for the outlier analysis.

Some examples of this effort can be seen below.



Approval Amount Outlier Analysis

Correlation heat-matrix to identify key dependencies

Heatmap of number of loans approved across all the states

Loan Approval Amounts Density Plot

Anomaly Detection

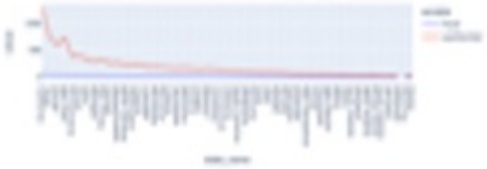
Standardization is done using standard scalar
Higher the anomaly score, higher the probability of loan being fraud

Isolation forest is being used to calculate anomaly scores
Anomaly score is calculated for features:
Initial Approval Amount, Payroll proceed and Jobs Reported

We then identify a loan being fraud if it lies in the top 95% of anomaly score calculated

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, Iapps to improve due diligence for debt-based Venture Capitalists

Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

Big Data Text Analysis - YouTube Channel Video Classification

[Project Slides] [DEMO]

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

Workflow

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

Sample Data Pipeline build on MAGE AI
Don't forget to get in touch with presenters for a live demo
Scan Down for a video Preview and codebase

Technology

- PYSPARK
- PYTHON
- FLASK
- PANDAS
- HTML/CSS
- JOBLIB
- LIGHTGBM
- BEAUTIFULSOUP
- SKLEARN
- XGBOOST
- WORDCLOUD
- MATPLOTLIB
- NLTK
- JOBLIB
- REGEX

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

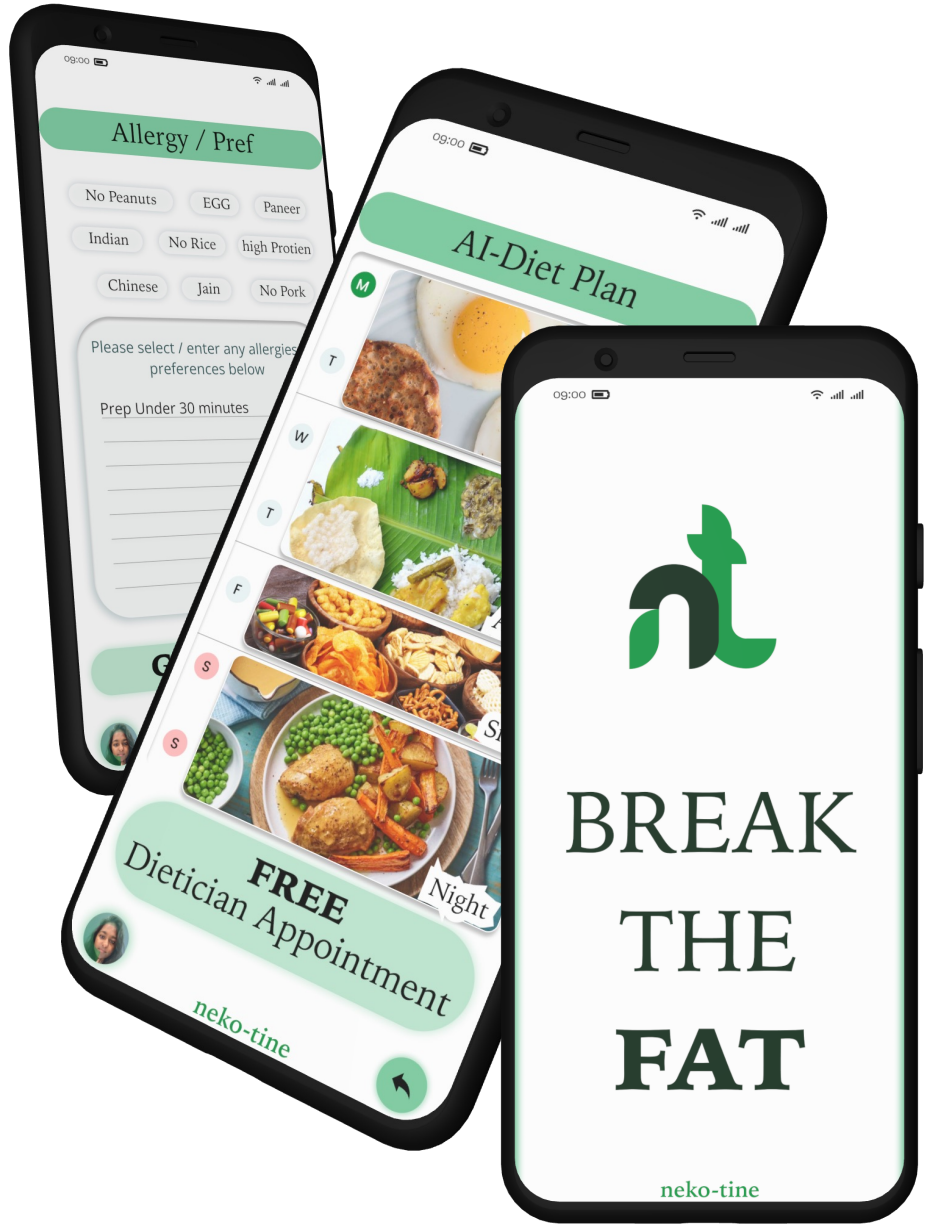
Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

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]



Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

Key Achievements

Developed Time Series Forecasting Model:

Implemented an efficient time series forecasting model using the AUTO ARIMA framework.

Achieved significant computation cost reduction and scalability for handling 10,000+ time series.

Implemented Recommendation System:

Developed a Collaborative Filtering and SVD's funk variant-based recommendation system.

Increased content engagement and monetization by recommending high-value unused keywords.

Streamlined Construction Project Tracking:

Strategized a real-time construction project tracking pipeline using data extraction from QuickBooks.

Generated insightful reports and visualizations using Excel and Tableau for effective project monitoring.

Automated Employee Transition Processes:

Automated data collection, storage, and task creation for employee transitions using Slack, ASANA, and Google Sheets.

Achieved significant time savings in administrative tasks related to equipment configuration.

Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

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. Amogh was skilled in utilizing Terraform, EC2, Lambda, and S3, which helped us to minimize costs and maximize efficiency. His experience in data engineering and data science also added to the overall success of the project. Amogh is a knowledgeable and dependable professional. He is well-suited for roles in software development, data engineering, and data science. I highly recommend Amogh to any employer looking for a skilled and reliable professional.

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. Amogh is a great problem solver and can come up with innovative solutions to complex problems. He is a great team player and can take the initiative in projects. He is an excellent communicator and can convey complex ideas to both technical and non-technical audiences. I highly recommend Amogh for software development, data engineering, and data science roles. He is the best person for the job and I am confident he will excel in any role he takes.

Referances available on request

Michael Washington
Director of Finance

RJ Lehman
Founder Stealth Startup

Andrew Jones
Director of IT

Dr Thomas Lechlar
Project Management Professional

Dr Joseph Morabito
Business Intelligence Professional

Dr Edward Stohr
BIA Program Director

Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari

Thank You

LINKEDIN.COM/IN/AMOGHKOKARI/

GITHUB.COM/AMOGHKOKARI

GETFOODRECIPE.STREAMLIT.APP

Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari