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 --> Data Engineer --> Data Scientist
Design Thinker + Project Manager + Analyst + Software Developer

Data Scientist | Python, SQL, Power BI, AI, Time Series & Predictive Analysis, Data Engineering |
I Help Businesses Mitigate Risk & Enhance Efficiency by Identifying Anomalies

aws certified
Cloud Practitioner
FOUNDATIONAL

Google | coursera
Google Project Management Certificate
Certificate of Completion

dbt
Learn
Fundamentals

Python

git

C3.ai

Streamlit

SQL

databricks

Gemini

LangChain

Azure

Excel

Power BI

pandas

Firebase

About Amogh

Hello, I'm an accomplished Analyst with passion for transforming complex datasets into actionable insights and strategic solutions across finance and various sectors. Holding a Master's in Information Systems, my expertise spans advanced statistical analysis, machine learning, AI technologies, and programming with a deep dive into **EXCEL**, **PYTHON**, **SQL**, **POWER BI** and more.

At the forefront of Data Science, I've engineered impactful models and systems, from **Forecasting** and **Recommendations** to optimizing operations and enhancing decision-making processes. My work with Con Edison and Bullwhip Technologies, among others, showcases my ability to leverage data for **predictive analytics**, **operational improvements**, and **insightful visualizations**, significantly affecting bottom lines and operational efficiency.

Certified in AWS Cloud, Google Business Intelligence, and as a Meta Database Engineer, I'm not just about data. I thrive in environments that challenge my problem-solving, agile methodologies, and stakeholder management skills. With a track record of leading projects from ideation through deployment, I excel in turning data into a compelling narrative for both technical and non-technical audiences.

Whether streamlining processes, automating data collection, or driving user-centered solutions, I'm dedicated to pushing the boundaries of what data can do to solve real-world problems, making data-driven decision-making the cornerstone of business strategy and innovation.

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

Big Data Text Analysis - YouTube Channel Video Classification

[Project Slides]   **[DEMO]**

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**

Workflow

Enter Youtube Channel URL

Channel URL [ ]

submit

Enter Youtube Channel URL

Channel URL [https://www.youtube.co]

submit

Enter Youtube Channel Title and Description

Channel Title
How to Make $10,000 in 24 Hours

Channel Description
make money online, how to invest, and more in this action-packed video. Plus, watch as MrBeast puts his money-making strategies to the test to see how much he can make in a single day!

Submit

Classifier Name	Value
GaussianNB	0
LGBMClassifier	1
XGBClassifier	0
AdaBoostClassifier	1

Word Clouds on all the video data present in the Channel

Description	Tags	Title

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
GBTClassifier	84.17	86.42	93.71	83.32
NaiveBayes	80.88	92.25	81.84	81.77

Run time Analysis

Normal (seconds)	Spark (seconds)
376.81	192.06

Data shape and class size

Rows x Cols	Lows (0)	Highs (1)
(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
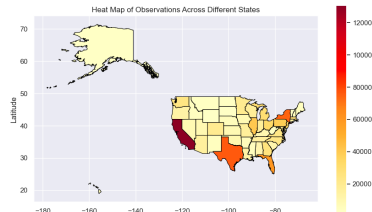
Amogh Mahadev Kokari -- linkedin.com/in/amoghkokari
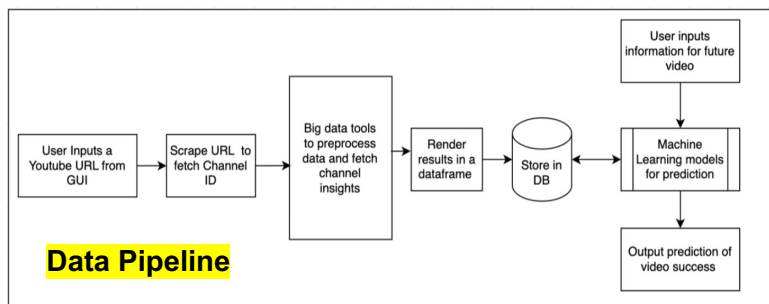
Fraud Detection -
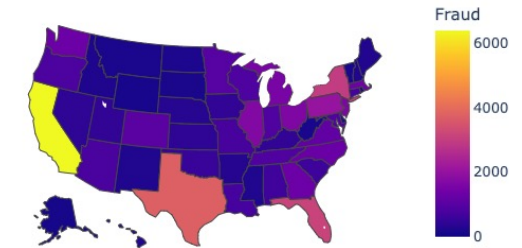# Paycheck Protection Program

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 Sklearn

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.
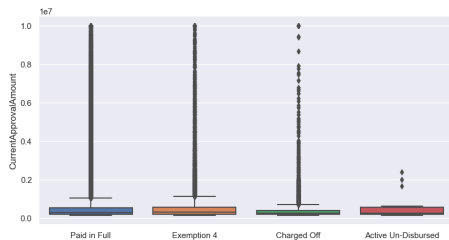
Approval Amount Outlier Analysis

Heatmap of number of loans approved across all the states

Correlation heat-matrix to identify key dependencies
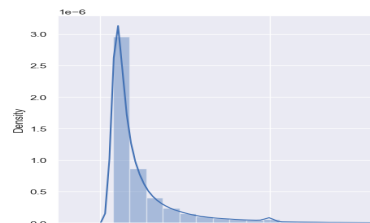
Loan Approval Amounts Density Plot

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)

State-wise probable frauds in USA
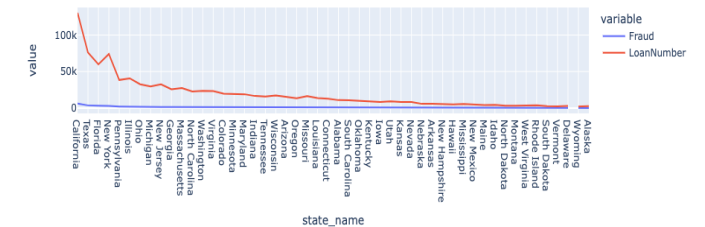
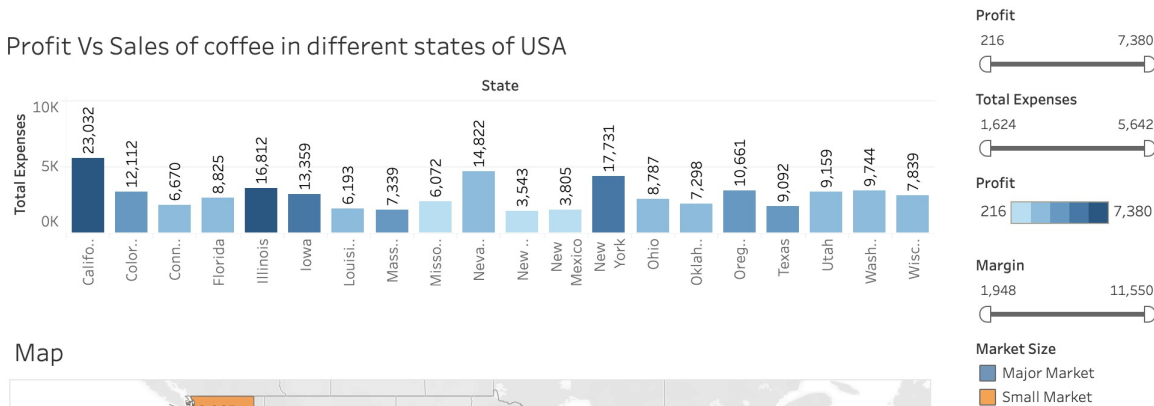State-wise Loans vs Frauds

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

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

Electric Vehicle Analysis [LINK]

Profit Vs Sales of coffee in different states of USA
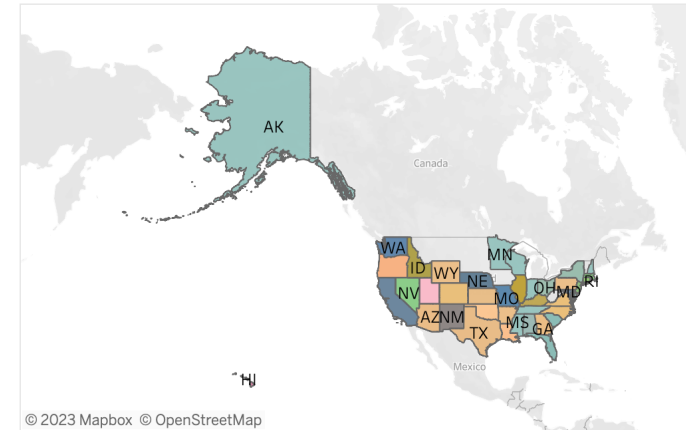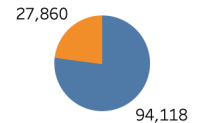
Profit
216 — 7,380

Total Expenses
1,624 — 5,642

Profit
216 — 7,380

Margin
1,948 — 11,550

Market Size
- Major Market
- Small Market

Map
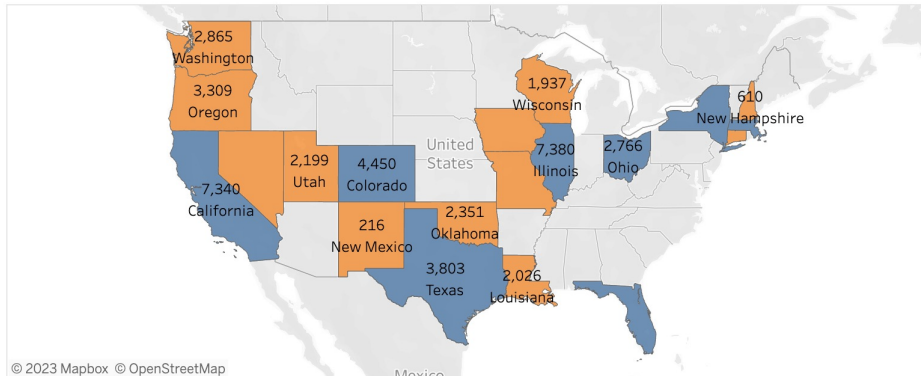
Electric Vehicle Population Data [Washington US]

Electric Vehicle Type
- Battery Electric Vehicle (BEV)
- Plug-in Hybrid Electric Vehicle (PHE...
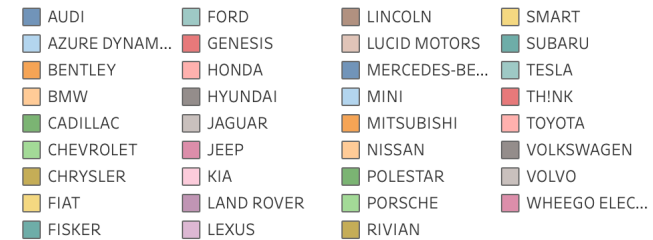
27,860 / 94,118

Clean Alternative Fuel Vehicle (CAFV) ...
- Clean Alternative Fuel Vehicle Eli...
- Eligibility unknown as battery ra...
- Not eligible due to low battery ra...

15,382 / 59,179 / 47,417

Make
- AUDI
- AZURE DYNAM...
- BENTLEY
- BMW
- CADILLAC
- CHEVROLET
- CHRYSLER
- FIAT
- FISKER
- FORD
- GENESIS
- HONDA
- HYUNDAI
- JAGUAR
- JEEP
- KIA
- LAND ROVER
- LEXUS
- LINCOLN
- LUCID MOTORS
- MERCEDES-BE...
- MINI
- MITSUBISHI
- NISSAN
- POLESTAR
- PORSCHE
- RIVIAN
- SMART
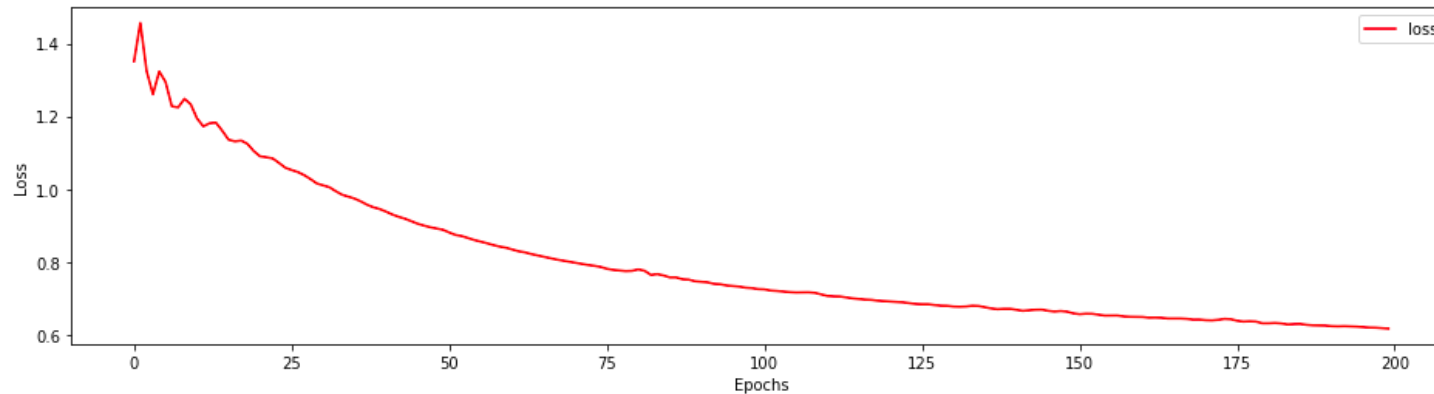- SUBARU
- TESLA
- TH!NK
- TOYOTA
- VOLKSWAGEN
- VOLVO
- WHEEGO ELEC...
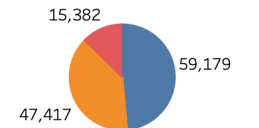
EEG Emotion Detection
Deep Learning Hard Parameter Sharing

Features range EEG.Cz - EEG.Fp2 for Disgust with intensity 3
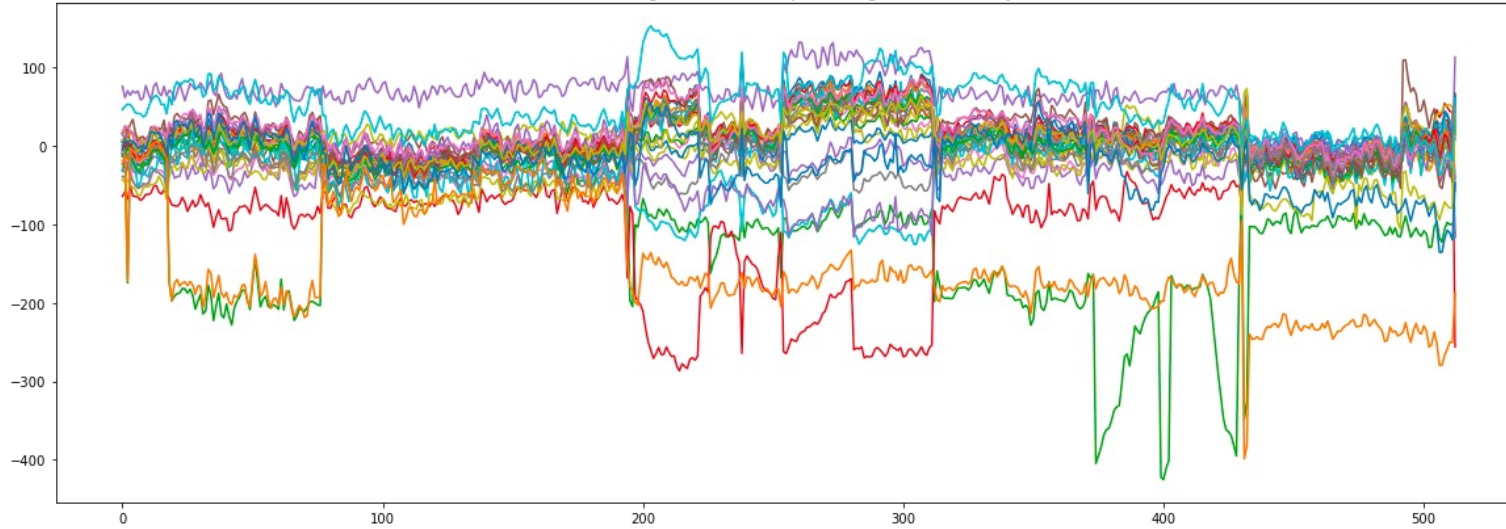
Total Loss

	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

Get Food Recipe [LINK] (LLM application) [code]   Get Fitness Routine [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 My Excercise 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/ plaace...

☁ 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)

1   10

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

How many days a week can you commit

1   7

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]

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)

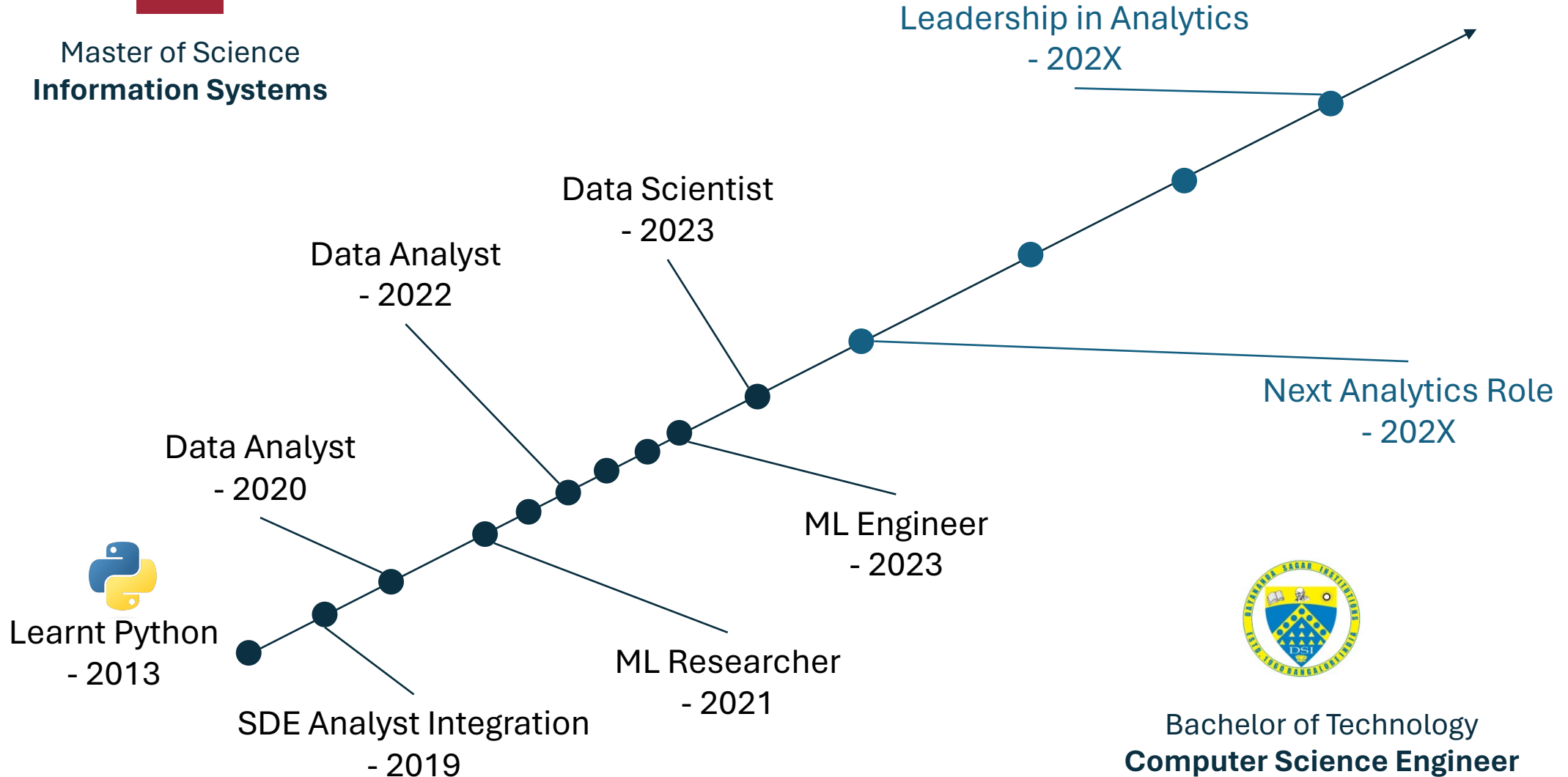[**Prototype**]   [**Project Slides**]

[**Explainer Video**]

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

CAREER TIMELINE

Master of Science
Information Systems

Leadership in Analytics
- 202X

Data Scientist
- 2023

Data Analyst
- 2022

Data Analyst
- 2020

Next Analytics Role
- 202X

ML Engineer
- 2023

Learnt Python
- 2013

ML Researcher
- 2021

SDE Analyst Integration
- 2019

Bachelor of Technology
Computer Science Engineer

9

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

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

Thank You

LINKEDIN.COM/IN/AMOGHKOKARI/

GITHUB.COM/AMOGHKOKARI

GETFOODRECIPE.STREAMLIT.APP