KIRUTHIKA. M

+91 8056242338 ◇ Tiruppur
kiruthika.m5825@gmail.com ◇ https://www.linkedin.com/in/kiruthika-m-55864737a/ ◇ https://github.com/Kiruthika5825
https://public.tableau.com/app/profile/kiruthika

EDUCATION

MBA Business Analytics and Applied Finance, RVS IMSR , Sulur — August 2022 - April 2024

BSC. Mathematics , St.Joseph's College for Women, Tiruppur — July 2019 - June 2022

PROFESSIONAL EXPERIENCE

CAPCHEK ANALYTICS — June 2024 - Present

- **Product & Analytics Lead**

  Designed a centralized analytics architecture by flattening buyer review data (expectation & post-delivery experience) from MongoDB → Supabase → Apache Superset, hosted and secured on AWS. Built dashboards covering product insights, support ticket performance, issue concentration, and resolution trends for brand teams to identify root problems faster.

  Defined customer trust and review KPIs with stakeholders and enforced Superset as the single source of truth for all analytical charts to ensure long-term maintainability and metric extensibility. Led API consumption strategy where developers fetch chart data from Superset for richer customer-facing visualizations, improving UI flexibility beyond native dashboards. Enabled smooth dashboard and chart embedding using Angular SDK & Angular SDK utilities in the Angular application.

- **Data Collection Workflow**

  Co-designed a user-friendly feedback escalation system that converts negative reviews into structured support tickets, routes brand responses through email-threaded chat views, and governs the ticket lifecycle to closure. Documented end-to-end flow plan to collect necessary data for furthur analytics and build dashboard using this data.

- **Leadership & Agile Delivery**

  Led a lean data team with ownership of task allocation, sprint planning, backlog refinement, and stakeholder alignment. Facilitated Scrum stand-ups and cross-functional syncs to remove blockers, enforce analytical validation, and deliver insights predictably for product and review systems.

ULTIWEB TECHNOLOGIES — Jun 2024 - Nov 2024

- As a Data Scientist intern, Contributed to a review intelligence platform, where I led initiatives on review analytics, customer sentiment tracking, and badge classification systems to boost brand-customer engagement.
- Designed pre- and post-delivery review surveys, action points, canned responses, and dashboards for customer and admin views.
- Built analytical frameworks to derive insights from customer reviews, including sentiment trends, feature mentions, and review-response impact.
- Worked for a client "DigitalBull Leads" automation workflows using SmythOS, streamlining video creation steps for digital marketing.

AFAME TECHNOLOGIES — May 2024

- The primary goal of this is to build and evaluate machine learning models that can accurately identify fraudulent transactions,
- In total of 983 unique credit cards 687 credit cards have fraud transactions (ie., 70 percentage of credit card holders face fraudulent transaction atleast once), which is a huge problem.

- Logistic regression performed poorly with zero precision and recall. Decision tree and random forest models outperformed logistic regression. Random forest achieved the highest precision(0.98) and accuracy(99.83) among the three models.

PROJECTS

RAG Pipeline: This project is a full-stack RAG system I developed to demonstrate real-world retrieval and LLM integration. It automates web scraping with BeautifulSoup, cleans and structures text into LangChain documents, and applies intelligent chunking for optimal context windows.

I generated embeddings using SentenceTransformers and stored them in a Milvus vector database with HNSW indexing for fast and scalable similarity search. The system exposes FastAPI endpoints for URL ingestion and question answering, enabling seamless data flow from raw webpages to LLM-powered responses.

I integrated NVIDIA NIM's Llama model through an OpenAI-compatible API to produce accurate, context-grounded answers. This project highlights my skills in building modular pipelines, designing scalable architectures, and working with modern AI infrastructure end to end.

Tech Stack: Python, BeautifulSoup, LangChain, Milvus, SentenceTransformers, FastAPI, NVIDIA NIM

Fake News Classification: Developed an end-to-end NLP pipeline following CRISP-DM methodology to classify 56,000+ news articles.Engineered features using TF-IDF vectorization and performed VADER sentiment analysis to identify key linguistic patterns in misinformation.

Achieved 99% accuracy using Logistic Regression, outperforming Random Forest and Decision Tree models optimized via cost-complexity pruning.

Stack Overflow Trend Analysis:

Data Extraction & Wrangling: Engineered a data collection pipeline using Web Scraping and APIs to harvest and clean records from 90,000+ global developers.

Market Intelligence Dashboard: Developed an interactive IBM Cognos Dashboard to visualize technology trends, identifying Python as the fastest-growing language and PostgreSQL as a rising database competitor.

Geospatial & Workforce Analytics: Analyzed job posting distributions to pinpoint Washington DC (5,300+ jobs) as the leading tech hub, correlating location data with developer demographics.

Tech Stack: Python, Numpy, Matplotlib, Seaborn, SQL, APIs, IBM Cognos, Web Scraping

SpaceX Falcon 9 Landing Prediction:

ETL Pipeline & Data Collection: Constructed a robust data pipeline extracting launch data via the SpaceX REST API and Web Scraping (BeautifulSoup) from Wikipedia, processing raw JSON into structured DataFrames.

Geospatial & Exploratory Analysis: Executed complex SQL queries to rank launch sites and utilized Folium to visualize geospatial proximity to railways and coastlines, identifying KSC LC-39A as the highest success rate site (76.9

Interactive Dashboarding: Built a dynamic Plotly Dash application to monitor launch outcomes, visualizing a positive correlation between lighter payloads (¡4000kg) and landing success.

Predictive Modeling: Optimized Classification models (Logistic Regression, SVM, Decision Tree) using GridSearchCV for hyperparameter tuning; identified the Decision Tree Classifier as the top-performing model for predicting first-stage landing success.

Tech stack: Webscraping, APIs, Python, SQL, Scikit-Learn, Plotly Dash, Folium.

SKILLS

Programming Languages	Python, R
Libraries	Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, NLTK, OpenCV, HuggingFace, Transformers
Databases	MySQL, Supabase
Business Intelligence	Power BI, Tableau, Excel, Apache Superset
Analytical Skills	Statistical Analysis, Data Science, Data-Driven Decision Making, Data Visualization, Feature Engineering, Stakeholder Collaboration, Competitive Analysis
Certifications	IBM Data Analyst Professional Certificate, IBM Data Science Professional Certificate, IBM Machine Learning