Anurag Mallick

Phone: +1 732-801-8730
Email: anurag.mallick@rutgers.edu
Linkedin: linkedin.com/in/anurag-mallick
Github: github.com/mallick20

EDUCATION

Masters in Data Science

- Rutgers, The State University of New Jersey 2024 - 2026

  Courses: Prob. & Statistical Inference for Data Science, Regression and Time Series, Data Structure and Algorithm,

B.E in Electrical and Electronics Engineering

- Birla Institute of Technology and Science, Pilani 2017 - 2021

WORK EXPERIENCE

- **Blackbox | Plano, Texas, USA** Dec 2024 - Present

  AI Engineer Intern

  - Support the development of AI-powered chatbot to improve access to SOWs, BOMs and Quotes by leveraging RAG techniques for efficient document retrieval.
  - Help ensure smooth implementation and takeover of AI solutions, while alos contributing to documentation and extending AI capabilities across use-cases.

- **Applied Data Finance | Chennai, India** Sep 2021 – Jul 2024

  Senior Data Scientist

  - Utilized NLP techniques to analyze and mitigate operational risks across various modes of communication. Developed models to identify risks with **70% precision** to make the process efficient.
  - Built Topic Models using BERTopic to segment customer requests. Effectively grouped **up to 80% of customer requests** into Valid Topics for more efficient resolution.
  - Led the foundation for improved resource management by categorizing and prioritizing cases for manual review, leading to more efficient review and response times.
  - Worked on finetuning LLM for better classification of customer complaints.

  Data Scientist

  - Built a classifier model using customer's repayment behavior to predict the success rate of payment retries and successfully segmented low-risk accounts, having a **37% higher success rate** than average.
  - Enhanced the Best Time to Call strategy to call customers according to their preferences and call history. Achieved a mean **increase of 20% RPCs** over six months.
  - Conducted in-depth analysis of returning customers to fix defects in the data and explore trends according to issues in disbursal plans.
  - Designed campaigns to engage delinquent customers and used statistical analysis to test the outcomes among various segments and improve collections.

RESEARCH EXPERIENCE

- **Real-Time Tweet Classification**

  BITS Pilani 2021

  - Implemented the foundational framework for classifying tweets based on their Non-Profit status.
  - Pre-processed raw tweets using regular expressions and developed the dataset processing pipeline.
  - Used BERTweet to build and integrate the classification pipeline's Level 1 (L1) and Level 2 (L2) stages.

- **Indian Language Machine Translation**

  IIT Roorkee 2020

  - Utilized deep learning framework to improve the accuracy of machine translation for Indian Languages.
  - Implemented word-level translation of OOV words, achieving a **BLEU Score of 73.5%.**

PROJECTS

- **Happiness Prediction**

  Course project [Regression and Time Series Analysis – Rutgers University] 2024

- As part of the course, participated in a competition to predict happiness index for different countries using predictors quantizing social and economic factors.
- Built a ensemble model using Bagging regression on Decision Trees as base estimators and determined important attributes influencing happiness.

- **Academic Assistant** [CODE] 2024
  - The project aims to develop a virtual assistant to streamline academic management. It will summarize and retrieve essential information, support academic tasks, and assist with course-related matters.
  - By using RAG techniques, the assistant is designed to provide answers relevant and tailored to course content.

WORKSHOPS

- RU Health Hackathon 2024
  - Implemented framework to streamline prior authorization process with AI-driven recommendations.
  - Designed a RAG based question-answering system to integrate history of the patient, insurance provider guidelines to provide feedback about prior authorization requirements reducing delays.

SKILLS

- **Data Analysis and Statistical Inference**
- **Machine Learning:** Linear Regression, Logistic Regression, Gradient Boosting, Bagging
- **NLP:** Document Classification, Topic Modelling, LLM Finetuning, RAG Application
- **ML Libraries:** Pytorch, Scikit-Learn, NLTK
- **Technical Skills:** Python, R, C++, SQL, Git, Microsoft Azure

AWARDS

- Secured 2nd position at ADF Data Science Hackathon – 2022
  Built an XGBoost Model to predict customer risk using Tradeline Attributes.