Piyush Jeena

Data Scientist

Data-driven researcher familiar with cleaning and modelling data for use by technical and non-technical personnel. Highly organized, motivated, and diligent with significant background in quantitative modelling, data science, and machine learning. Looking to work in an organisation that provides opportunities for technical and personal advancement.

jeenapiyush101@gmail.com

+33 0780745677

Bordeaux, France

pjeena.github.io/devportfolio/

linkedin.com/in/piyush-jeena

github.com/pjeena

WORK EXPERIENCE

PhD Researcher in Computational Physics

CNRS

10/2019 - 11/2022 *Bordeaux, France*

Achievements/Tasks

- Developed state-of-the-art Markov Chain Monte Carlo simulations in Python from scratch to generate quantitative diagrams for real world materials and achieved > 90% accuracy against experimental results.
- Improved the baseline model by implementing Parallel Tempering and reduced the simulation time by a 100x.
- Utilized visualization, analytical and statistical techniques to analyze and process complex data by launching jobs using Slurm in a High-Performance Cluster.
- Streamlined research processes to meet tight deadlines for multiple projects and wrote research papers, reports, dissertation and presented work in multiple prestigious conferences.

PERSONAL PROJECTS

Podcast recommender system

- Designed and developed an end-to-end machine learning pipeline where the user chooses a podcast listened to before or enters a general description and the model recommends options from a database of 2300+ podcasts with 19 categories and 500k+ reviews. The data was aggregated from itunes and Spotify.
- Cleaned and preprocessed the raw data and generated word embeddings for all the podcasts using TF-IDF and all-MiniLM sentence-transformer models. The podcasts were recommended based on the cosine-similarity scores between the input query and the database.
- Performed one-tailed t-test to examine whether the between-category scores were significantly lower than the within-category scores and achieved 95% accuracy.
- Built a web UI and deployed it on AWS EC2 instance.

VISA Profiling of employees

- Developed a data-driven solution to facilitate VISA approvals and reduce the time spent in the process.
- Performed preprocessing and feature engineering to improve the data quality and implemented various classification models (Logistic Regression, Random Forest, Bagging and Boosting classifiers) to predict the visa status.
- Achieved a F1-score of 82% and increased the accuracy by 10% from the baseline by tuning the hyperparameters of XGBoost model.
- Utilized MLflow to manage and track experiments to record and compare parameters and results.

Data Engineering pipeline to visualize food recipes

- Scraped 35000+ recipes from Allrecipes using Python and leveraged Airflow to monitor workflow on AWS EC2 instance. Preprocessed the data and launched AWS Glue job to migrate data from S3 to Redshift.
- Wrote SQL queries on Redshift and visualized it on Amazon Quicksight.

TECHNICAL SKILLS

Programming Languages
Python, SQL, C.

Python Packages and Framework
NumPy, SciPy, Matplotlib, Seaborn, Pandas, Scikit-learn, PyTorch, NLTK.

Cloud Deployment and frameworks
AWS EC2, Streamlit, Git, Airflow.

MLops Tools
MLflow.

Data Science/ Machine Learning
Data Visualization, Data Preparation, Feature Engineering, Statistical Modelling, Quantitative Analysis, Natural Language processing(NLP) etc.

SOFT SKILLS

Problem Solving | Critical Thinking | Adaptability | Curiosity | Communication | Time-Management | Storytelling

EDUCATION

PhD in Physics

CNRS and University of Bordeaux

10/2019 - 12/2022 *Bordeaux, France*

MS in Physics

IIT Bombay

07/2017 - 07/2019 *Mumbai, India*

LANGUAGES

English
Full Professional Proficiency

Hindi
Native or Bilingual Proficiency

French
Limited Working Proficiency