Anshika Gupta

Email: anshikagupta379@gmail.com | My LinkedIn | Portfolio | Github | Mobile: +91 8824623961

Skills

Programming Languages: Python, Java, C++, SQL, HTML/CSS/JavaScript
Frameworks: Scikit-Learn, BeautifulSoup, Keras, TensorFlow, PyTorch, PySpark, Pandas, Numpy, Matplotlib, SciPy, Flask, Django, FastAPI, Apache Hadoop, Spark
Tools: MySQL, Oracle, MongoDB, SAS, Tableau, Power BI, Excel VBA, Google Analytics, AWS (S3, SageMaker, Redshift, Lambda), Microsoft Azure, Snowflake, Informatica, Google Cloud Platform, KNIME, Docker, Git
Miscellaneous: Statistical Methodologies, Handling datasets with missing values and class imbalance, Time Series analysis, Predictive modelling, Optimizing model performance, Natural Language Processing (NLP)

Work Experience

Mazars India Gurugram, India
Data Analyst Intern Feb 2023 – Apr 2023

- Extracted relevant financial information from sources like ERP systems and CRM databases using complex SQL queries for auditing purposes.
- Coordinated with seniors to implement Excel Macros for data cleaning and data standardization tasks improving efficiency and reducing time and error by up to 30%.
- Employed DAX within Power BI to create custom calculations, such as financial ratios, KPIs and trend analysis.
- Led an intern team in developing and deploying a machine learning model for automating anomaly detection in audit data. (Python, FastAPI), made auditing significantly faster.
- Implemented performance-critical modules in C++ for an anomaly detection system, optimizing data processing speed and memory usage by 15%.
- Compiled findings, reviewed results, and shared insights with stakeholders using Tableau dashboards and reports.
- Executed web scraping tasks for competitor analysis (BeautifulSoup library).

UpSkillz Gurugram, India
Machine Learning Researcher Jul 2022 – Sep 2022

- Performed Exploratory Data Analysis (EDA) on highly imbalanced datasets such as cash fraud detection (improved recall from 0.82 to 0.91).
- Optimized performance of Kaggle's Tumor classification models through Transfer Learning, tuning parameters, and modifying layers to improve model accuracy. (AlexNet, GoogleNet, ResNet, VGG16). Presented an in-depth review and recommendation of 10 such deep learning models.

Education

The NorthCap University, Gurugram, India Oct 2020 – Jun 2024
B. Tech in Computer Science (Specialization: Data Science) **CGPA**: **8.75/10**
Relevant Coursework: AI and ML, Deep Learning (CNN, RNN, LSTM, BERT), NLP, Data Engineering, Data Mining, Applied Statistics, Business Intelligence and Data Visualization, Database Management System, OOP, Data Structures and Algorithms, Linear Algebra, Differential Equations, Computer Networks, Software Engineering, Operating Systems.

Relevant Projects

1. Crop Disease Prediction using Deep Learning (97% Accuracy): Convolutional Neural Network (CNN) model to predict crop disease in an uploaded image (14 plants, 26 diseases), recommend treatment, developed using TensorFlow, Keras, compared performance to popular architectures by fine-tuning them. Deployed by FastAPI, Postman, front-end - HTML/CSS and JS.
2. Customer Segmentation and Recommendation System: A transactional database (541,908 entries), performed EDA, PCA, analyzed customer behavior and purchasing patterns. Built a recommendation system using collaborative and content-based filtering. Used K-means clustering based on Recency, Frequency, Monetary (RFM), effectiveness studied using silhouette scores, elbow method, radar charts.
3. Health Insurance Claim Analysis: Predictive regression modeling for estimating health insurance claims, handled null values, captured trends by pyplot, tested for difference in mean claims in different age groups using ANOVA. Used Linear Regression, Decision Tree, Random Forest, Gradient Boosting, XG Boost, Light GBM, evaluated by MSE and R2 score.
4. Credit Card Fraud Detection Model (99% Accuracy): PCA-transformed dataset (284,807 transactions), class imbalance corrected using Borderline SMOTE, t-test to compare difference in mean amounts of both classes. Algos - Logistic Regression,

Naïve Bayes, KNN and Random Forest. Evaluation metrics - Accuracy, Precision, Recall, F1, AUC score and z-test to check for significant statistical difference in models' performance.

5. Power BI Analysis of IPL Performance and Trends: Descriptive analysis of IPL data of the decade, included DAX for custom calculations, filters, slicers and pivots creating charts, insights, dashboard and plots. Provided prescriptive insights.

6. Flight Delay Analysis and Visualization using Tableau: Analyzed factors contributing to flight delays, cancellations, and diversions across various airports and airlines, created interactive dashboards including charts, trends, and use of Calculated Fields.

Accepted Publications

Gupta, A., Solanki, D. (2023). Crop Disease Prediction using Deep Learning. Accepted by the International Journal of Innovative Science and Research (IJISRT).

Certifications

- Data Structures and Algorithms: Coding Ninjas
- Top Performer in DSA: Coding Ninjas
- Data Analytics: ICT Honeywell
- Machine Learning Foundations: AWS
- Cloud Foundations: AWS