Mobile: 07746740962
ishitavidhi@gmail.com
GitHub repository
LinkedIn

ISHITA AGARWAL

Data analyst in the financial industry with 4 years of motivated Big Data experience, specializing in translating complex technological requirements into clear and compelling narratives for non-technical stakeholders. Proficient in Azure Cloud Services and AWS, with a solid foundation in cloud computing principles. Actively expanding skills in DevOps, utilizing tools like Docker, Jenkins, and AWS CodeDeploy to enhance operational workflows. Knowledgeable in testing and assurance processes, with a firm grasp on reference data management. **On the administrative side, I am currently on PSW and don't require sponsorship until January 2026.**

EDUCATION

MSc Data Science, University of Essex, UK — **October 2022 – 2023**

- Relevant Modules: Modeling Experimental Data, Machine Learning, Text Analytics, Neural Networks and Deep learning, Databases and Data processing with SQL, Data Visualization with R, Applied Statistics, Probability
- **Grade**: Distinction

BTech Information Telecommunications, SRM University, India — **August 2014 - May 2018**

- **Grade**:8.972

SKILLS

- **Tools:** Cloudera, Pepperdata, WXM, Tableau, Looker, Microsoft Excel, (Pivot) Microsoft Word, PowerPoint, Microsoft PowerBI, ETL
- **Programming:** Python (Pandas, NumPy, PyTorch, Scikit-learn, Tensor Flow, NLTK, Spacy), R, C++, Java
- **Big Data Technologies:** Hadoop Stack (Map Reduce, Apache Spark, Hive, Impala, Advanced SQL, MS SQL)
- **Operating System:** Linux
- **Data Analysis & Modelling:** Machine Learning (ML), Deep Learning, Natural Language Processing (NLP), Hugging Face Transformers for advanced model implementation
- **Agile Methodologies & Project Management Tools:** AGILE, Jira, Confluence, Git Hub
- **Soft Skills**: Problem Solving, Strong communication, Collaborative Teamwork, Technical Writing, Presentation, Quantitative thinking
- **Certification**: AZURE Fundamentals, AZURE AI Fundamentals Certified

WORK EXPERIENCE

August 2018 – September 2022

Data Analyst, Bank of America

- Experienced in developing, extracting, loading and executing queries on **complex databases and datasets using SQL, Hive, Impala.**
- Skilled in ensuring data integrity and data security, managing user permissions, and conducting data quality checks to maintain accuracy.
- **Summarized findings through high-quality visualizations and reports in Tableau and Excel to analyze key metrics, increasing lead forecasting accuracy by 25% and reducing decision-making time by 50%.**
- Closely engaged with application owners and stakeholders to understand, manipulate and transform **intricate business challenges, delivering comprehensive analyses that improved job performance by 40%.**
- Maintained complex Excel spreadsheets to track critical metrics, fostering **strong internal and external stakeholder relationships to support reporting activities and resolve data issues, demonstrating keen numeracy skills and attention to detail.**
- Automated various processes using **Python and API integration, achieving a 30% reduction in operational costs** and a significant decrease in manual hours.
- Experience in Power BI Report builder experience (including DAX)
- Familiar with Agile methodologies and used tools **like JIRA, Confluence, and GitHub for project tracking and version control.**
- Demonstrated the ability **to draw process maps, write user stories, and epics,** essential for effective Agile project management.

Technical Support Engineer, Bank of America

- Experience in **monitoring performance of all hadoop 10+ production clusters** from UAT to production of bank and **automated** various process **using Python to increase operational efficiency and reduce costs**.
- **Strong ability to troubleshoot** and resolve hadoop platform issues of 6000+ servers and managing the risk.
- Offered quickly implementable technical solution to **optimize hadoop performance** managing memory and core by identifying resource consumption services for various production jobs in the Spark/Map Reduce ecosystem to maximise hadoop performance.
- Collaborated with stakeholders of various applications to provide in-depth analyses of the causes of production problems, offer recommendations for the best fixes, and express opinions on challenger models.
- Solid understanding of performance tuning tools like Cloudera, WXM and Pepperdata to find practical business solutions.
- Managed documentation efforts, detailing processes and solutions to address targeted challenges and achieve business goals.

PROJECT

Text Classification and Offensive Speech Detection

- Designed, implemented, and evaluated a text classifier to identify offensive speech on OLID dataset, employing state-of-the-art natural language processing techniques to eliminate regular expressions, emoticons, email addresses, and other irrelevant words and further partitioned the data into four equal subsets, using each to train the models.
- Deployed a text classification model using deep learning architectures CNN and machine learning model Logistic Regression and in Python (TensorFlow, scikit-learn achieving accurate identification of offensive and non-offensive text.
- Achieved superior performance with the SVM and TF-IDF Vectorizer model, attaining an 80.47% accuracy rate.

Project on Optimizing User Engagement

- Designed and executed an A/B test to evaluate the effectiveness of two different user engagement strategies on a web platform.
- Analyzed user interaction data using Python and statistical methods to determine the superior strategy. The successful strategy led to a 20% increase in user engagement metrics.

House Price Prediction and Market Analysis

- Utilized statistical and machine learning methods to analyze the "house data.csv" dataset as a data analyst at an estate agency.
- Identified and analyzed patterns and trends in the housing market through data interpretation.
- Conducted exploratory data analysis to understand the dataset's structure and characteristics.
- Utilized graphical representations to visually explore data patterns, outliers, and variable relationships.
- training.

Real-Time Fraud Detection in Payment Systems

- Developed and implemented a machine learning model using Python, TensorFlow, and Scikit-Learn to detect fraud in payment systems, focusing on real-time analysis of transaction data to enhance security and reduce fraud.
- Conducted feature engineering to identify critical fraud patterns and applied a mix of algorithms such as decision trees, Random Forest and LSTM networks, tailored for the sequential nature of transaction data.
- The model significantly improved fraud detection accuracy, leading to a notable reduction in financial losses and bolstering transaction security.

Text-to-Image Generation Using Stable Diffusion

- Developed an interactive web application on Python leveraging cutting-edge AI technology to transform textual descriptions into vivid images.
- The application utilizes different Stable Diffusion models from Hugging Face's diffusers library, integrated through a Streamlit interface, to allow real-time user interaction.