Kashif Mahmood

Data Scientist

Kashif03371733@gmail.com | +92 3371733897 | LinkedIn | Portfolio | Github

ABOUT ME

I'm Kashif, an aspiring Data Scientist with over one year of hands-on project experience. I have developed several data-driven projects that demonstrate my strong technical abilities and problem-solving skills. For instance, I worked on a pipeline condition prediction project and achieved 14% improvement in accuracy over the historical benchmark on Kaggle by building a model well-suited to the problem. My expertise spans databases, statistics, and core data science tasks such as exploratory data analysis and feature engineering, alongside model development using machine learning and deep learning techniques.

EDUCATION

University of Sargodha

BS Software Engineering (3.90 / 4.00 CGPA) 2021-2025

- Excellent in programming and problem-solving coursework.
- Led a high-impact Final Year Project on Peanut Classification, Pest Identification, and Disease Detection, and set new benchmarks with over 98% accuracy. Research papers are currently under review in IEEE and PLOS ONE.
- Served as Head of the Data Science Group, where I organized events focused on emerging trends and data- driven approaches.

SKILLS

- **Data Science & Machine Learning:** Python, Statistics, Database, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, PyTorch, BeautifulSoup
- **Tools & Technologies:** Jupyter Notebook, Power BI, MySQL Workbench, Git & GitHub, JIRA
- **Core Competencies:** Leadership, Communication, Critical Thinking, Problem Solving

PROJECTS

PeanutInfo - Smart Agriculture System

- Constructed a custom peanut classification dataset with 1700+ images and integrated pest and disease datasets from Kaggle, based on baseline research papers to support accurate variety classification, disease detection, and pest identification.
- Employed 7 state-of-the-art Pre-trained models including ResNet-50, ResNet-101, EfficientNet-B0/B4, and ConvNeXt-Tiny, Small, and Base, achieving over 98% accuracy and setting new global benchmarks in peanut image classification.
- Designed and developed a full-stack, real-time prediction platform using HTML, CSS, Bootstrap, PHP, MySQL, and Flask API, enabling image uploads and seamless interaction with deployed deep learning models.

PipeSense - Pipeline Condition Prediction System

- Built a robust regression model using a Kaggle dataset to predict pipe conditions based on physical and environmental features, including Pipe Size, Diameter, Thickness, Material, Strength (MPa), Maximum Pressure, Material Loss %, Thickness Loss, and Temperature.
- Applied Yeo-Johnson transformation to normalize skewed features and used StandardScaler for consistent feature scaling. Handled categorical data using Ordinal Encoding for ordered variables and One-Hot Encoding for nominal variables.
- Automated the end-to-end ML workflow using Pipeline and ColumnTransformer. Achieved 98.1% accuracy, a 14% improvement over the historical Kaggle benchmark. Deployed the model with real-time compatibility using Pickle.

Employee Retention Predictor – ML-Based HR Analytics Tool

- Built an end-to-end classification model using HR data (12+ features, 14,999 records) to predict employee attrition based on working hours, salary levels, tenure, and more.
- Performed data cleaning, EDA, feature scaling, and applied One-Hot Encoding for categorical variables. Split the dataset for training and validation using 80/20 train-test ratio.
- Trained and evaluated models including Logistic Regression and Decision Tree, achieving an accuracy of 80% and assessing performance using accuracy scores and the confusion matrix.