

M. AbuBakar Arshad

(Data Scientist)

Email Id: abubakr510@gmail.com

Contact No: +92 310 4416466

LinkedIn: <https://www.linkedin.com/in/rao-abubakar-710b231b3>

Github: <https://github.com/MAbubakarArshad>

Work Experience

Algo as ETL Developer intern (July 2022 – Nov 2022)

- Azure Data Factory, Azure Synapse Analytics & SSIS for ETL SQL Server for Managing, modifying & storing the Data.
- Using Power BI for Reports & Dashboards including DAX.
- Implementation of Azure Analysis Service Cube & used for reports in MS Power BI.

Term Projects

Ocular Disease Prediction (FYP)

The development of an ocular disease recognition system begins with the assembly of a diverse dataset containing eye images and pertinent data. This dataset undergoes preprocessing, involving tasks such as image resizing and numerical normalization, to ready it for analysis. Subsequently, deep learning or machine learning models are employed to extract meaningful features from the data, which forms the foundation for training and evaluation.

Following feature extraction, the chosen model is trained using the preprocessed data. This training process entails finetuning the model's parameters to minimize prediction errors. The model's effectiveness is then assessed through rigorous evaluation using a separate test dataset. Upon achieving satisfactory performance, the model is seamlessly integrated into an intuitive user interface, ensuring easy interaction for medical professionals.

A comprehensive testing phase ensues to validate the system's functionality and reliability, mitigating potential issues before deployment. Once deemed robust, the ocular disease recognition system is rolled out for real-world utilization. Its capability to accurately diagnose eye conditions aids medical practitioners in providing timely and precise treatments. Continual updates and vigilant monitoring sustain the system's efficiency, making it a valuable asset in enhancing patient care and diagnostic accuracy.

Credit Card Fraud Detection

The initial step in preparing credit card transaction data for analysis involves rigorous pre-processing to guarantee data quality, consistency, and suitability. This entails removing any null values to eliminate gaps in information and ensure that subsequent analysis remains reliable. To facilitate the application of machine learning algorithms such as SVM, Random Forest, and KNN, all data is transformed into numerical format, enabling seamless compatibility.

The next stage addresses potential outliers and noise present within the data. These anomalies can distort analysis and model performance, necessitating careful handling. By identifying and managing outliers and noise, the accuracy and effectiveness

of the ensuing machine learning models are preserved. Additionally, data normalization or scaling is undertaken to establish a level playing field for comparison among features, promoting optimal model performance across various algorithms.

Feature engineering and selection further enhance the prepared dataset. This step involves extracting meaningful and pertinent features that hold significant value for fraud detection. By curating the dataset to retain only relevant information, the subsequent machine learning models can better discern patterns and anomalies indicative of fraudulent transactions.

A paramount concern throughout this process is maintaining data integrity and privacy. Robust measures for data anonymization and security are incorporated to safeguard sensitive information. This proactive approach ensures that the pre-processed credit card transaction data is not only well-prepared for analysis but also respects confidentiality and compliance standards, thereby laying the groundwork for effective fraud detection models.

House Price Prediction Using Machine Learning

To embark on the task of predicting house prices through machine learning, the process commences with the assembly of a comprehensive dataset, encompassing pertinent details about houses including factors like size, location, room count, amenities, and more. The subsequent step involves meticulous preprocessing of the acquired data. This entails addressing missing values, eliminating outliers, and standardizing numerical attributes. For categorical variables, adept techniques such as one-hot encoding are employed to facilitate effective utilization within the model.

The journey towards accurate predictions entails the strategic selection of influential features capable of significantly contributing to the forecasted house prices. Techniques such as correlation analysis or feature importance ranking are harnessed to discern these pivotal attributes. A fitting regression algorithm is then chosen to suit the dataset's intricacies. Widely recognized options include linear regression, decision trees, and the robust random forests algorithm. The dataset is judiciously divided into training and testing subsets, ensuring a balanced representation of information.

With the groundwork set, the training data is input into the selected model, enabling it to discern intricate patterns and correlations between features and the target variable – house prices. Crucially, the model's efficacy is gauged through the prism of key evaluation metrics like mean squared error (MSE), root means squared error (RMSE), mean absolute error (MAE), and the R-square value. This rigorous assessment unveils the model's performance and its capacity to furnish accurate house price predictions, shaping the pathway toward optimal predictive capabilities.

Courses and Certificate

1. SQL for Data Science: This course was designed to give me a primer in the fundamentals of SQL and working with data so that I can analyze it for data science purposes. It provided the foundation and gradually makes you write both simple and complex queries to help you select data from tables. I started working with different types of data like strings and numbers and discuss methods to filter and pare down my results.
2. Machine Learning Specialization: Through a series of practical case studies, I could apply experience in major areas of Machine Learning including Prediction, Classification, Clustering, and Information Retrieval. I learned to analyze large and complex datasets, create systems that adapt and improve over time, and build intelligent applications that can make predictions from data.

3. Machine Learning with Python: This course began with a gentle introduction to Machine Learning and what it is, with topics like supervised vs unsupervised learning, linear & non-linear regression, simple regression, and more. I dive into classification techniques using different classification algorithms, namely K-Nearest Neighbors (KNN), decision trees, and Logistic Regression. I also learn about the importance and different types of clustering such as k-means, and hierarchical clustering.
4. Build Dashboards in Power BI: Through this project, I created a Dashboard in Power BI. I get data to bring into a model, build several reports, generate informative charts from each report, then choose powerful visuals to highlight on a Dashboard.

Skills

- Proficient in all major data science software including Azure, Excel, Power BI, Dax Studio.
- Strong base of knowledge and skill in Major machine learning and deep learning libraries, Pandas, NumPy, Scikit-learn, TensorFlow, and Keras.
- Top written and oral communication skills allowing for more productive conversations with clients and colleagues.
- Strong Background in Data Science Software like Azure, Power Bi and working with distinctive libraries.
- Motivated team player with excellent public speaking ability and oral, written, and Interpersonal communication.
- Initiator and problem-solver using creativity, resourcefulness, and assets to break down and overcome organizational obstacles.
- Lifelong learner committed to staying on top of the best current and emerging data science concepts and machine learning, especially in an international context.

Technical Skills

- Machine Learning
- Deep Learning
- Power BI
- SQL Server
- Python
- ETL
- DAX
- Data Warehouse
- Data Mining

Achievement

Ngiri Scholarship Recipient, [July 2023]

Awarded the Ngiri Scholarship for outstanding performance in FYP.

Education

2019-2023: Bachelor of Science in Data Sciences (University of Management and Technology, Lahore)