



SHOAIB HOQUE

CERTIFIED DATA SCIENTIST

PROFILE

With a solid foundation in Python programming, data analysis, data processing, and model building, I possess a Post Graduate Diploma in Data Science and a B.Sc. in Computer Science Engineering. My expertise lies in a profound comprehension of machine learning algorithms, creating data-driven stories, and data handling. I am skilled in harnessing Python libraries and machine learning resources to convert intricate data into practical insights. Through the achievement of the IABAC certified Data Scientist Certification, Google Advanced Data Analytics Professional Certificate and IBM Data Science Professional Certificate, my dedication to ongoing learning and involvement in cutting-edge data science and machine learning endeavors is unwavering.

WORK EXPERIENCE

Somikoron Ai

Jan 2024 - June 2024

Data Science Intern

- Extract crucial information by utilizing web scraping tools such as "Beautiful Soup" and "Selenium."
- Employ the acquired data to develop a "News Alert API Handler" that integrates NLP and other methodologies.

Rubixe

Data Science Consultant

Oct 2023 - May 2024

- Collaborated with the team members to understand client requirements and business challenges. Worked on a client project where machine learning algorithms like decision tree, XGBoost, random forest algorithm were applied and achieved 95% test accuracy. Played a key role in the project team and ensured timely submission of the projects.
- Constructed predictive models using machine learning algorithms, including Logistic Regression and Multiple Linear Regression, achieving an accuracy rate of 85% in attributing revenue and conversions to specific marketing touchpoints

Nogor Solutions Limited

May 2023 - Aug 2023

Data Analytics Trainee

- Conduct statistical analysis using Python scripts exclusively, without depending on external libraries.
- Assisted in data collection, data cleaning, and data preparation for analysis, ensuring data integrity and accuracy throughout the process.
- Conducted statistical analysis and hypothesis testing to validate business assumptions and uncover insights, resulting in data-driven process improvements.

CONTACT

+8801681855181

shoaibhoque@gmail.com

Dattapara, Tongi, Gazipur



EDUCATION

Feb 2023 - Oct 2023
UNITED INTERNATIONAL
UNIVERSITY

- Post Graduate Diploma in Data Science
- CGPA - 3.95 / 4.00

May 2018 - Dec 2022
SHANTO MARIAM
UNIVERSITY OF CREATIVE
TECHNOLOGY

- B.Sc. in Computer Science Engineering (CSE)
- CGPA - 3.73 / 4.00

Jan 2014 - Dec 2017
ISLAMIC UNIVERSITY
KUSHTIA

- BA. in Islamic Economics
- CGPA - 3.50 / 5.00

LANGUAGES

- English (Fluent)
- Bangla (Fluent)
- Arabic (Fluent)
- Spanish (Beginner)

TECHNICAL SKILLS

- **Programming Languages:**
Python, SQL
- **Machine Learning:**
Regression, Classification, Clustering, Time Series Analysis
- **Data Visualization:**
Matplotlib, Seaborn, Tableau, Power BI
- **Tools:** Pandas, NumPy, Scikit-learn, TensorFlow, Keras, Sweetviz, Autoviz, SMOTE
- **Web Scrapping:**
BeautifulSoup, Requests, Selenium
- **Statistical Analysis:**
Hypothesis Testing, A/B Testing, Kurtosis, Skewness, Distribution testing
- **Database Management:** MS SQL Server, MySQL, MongoDB
- **Version Control:** Git
- Data Cleaning and Pre-Processing
- Exploratory Data Analysis
- Natural Language Processing

SOFT SKILLS

- Project Management
- Adaptability
- Teamwork
- Time Management
- Leadership
- Effective Communication and Presentation Skills
- Critical, Strategic, Analytical Thinking
- Decision Making
- Storytelling

CERTIFICATION

- [IABAC Certified Data Scientist](#)
- [IBM Data Science Professional Certificate](#)
- [Google Advance Data Analytics Professional Certificate](#)
- [Datamites Certified Data Scientist](#)

PROJECTS

• **Improve ITSM processes with Machine Learning** | [Project Link](#)

Client: ABC Tech (Rubixe project)

My Contribution:

- Predicted High Priority Tickets: To predict priority 1 & 2 tickets, so that they can take preventive measures or fix the problem before it surfaces
- Applied Random Forest, Decision Tree, KNN, SVM, Gradient Boosting, Logistic Regression and Naive Bayes algorithms also used Hyperparameter Tuning of Grid Search CV and Randomized Search CV
- The Random Forest Classifier outperforms other models in predicting high priority tickets, exhibiting high accuracy of 89%. Logistic Regression and Naive Bayes models show lower performance, indicating the need for improvement. Random Forest, Decision Tree, K Nearest Neighbors, Support Vector Machine models perform competitively, offering viable options. The choice of the most suitable model should consider a balance between accuracy, precision, and recall, with further refinement recommended for practical applicability in addressing the priority ticket prediction problem.

• **Employee Performance Analysis** | [Project Link](#)

Client: INX Future Inc. (IABAC Project)

My Contribution:

- Found the required insights below:
 1. Department wise performances
 2. Top 3 Important Factors affecting employee performance
 3. A trained model which can predict the employee performance based on factors as inputs. This will be used to hire employees
 4. Recommendations to improve the employee performance based on insights from analysis.
- Completed with Data Preprocessing, Feature Selection, EDA and experimented with seven different models and Hyperparameter Tuning
- From predicting with multiple models, Random Forest and Gradient Boosting models are giving maximum accuracy level with 93.75% and 92.92%. Naive Bayes is producing lowest accuracy score with 74.58%. Logistic Regression and SVM models are producing moderate level accuracy among others with 85% and 84.17%.

• **Falcon 9 First Stage Landing Prediction** | [Project Link](#)

Client: Space X (IBM Project)

My Contribution:

- Collected data with SpaceX API, Web scraping launches records from Wikipedia, predicted with 4 different ML models and got 83.33% accuracy, did launch site's locations analysis with Folium also did EDA and Feature Engineering, data preprocessing.