

Muhammad Ihtisham Ul Haq

Data Scientist / NLP Engineer



ahtiisham.maliik@gmail.com



<https://www.linkedin.com/in/ahtisham-malik/>



Lahore



<https://github.com/ahtisham-maliik>



+92-303-9229203

Data Scientist with hands-on experience in feature engineering, data visualization, and predictive modeling. Transitioning from a software engineering background with strong proficiency in Python, SQL, and analytical problem-solving. Skilled in applying machine learning and NLP techniques to real-world datasets, extracting actionable insights and building robust models. Proven ability to deliver results through end-to-end data science projects, and eager to contribute to data-driven decision-making in a collaborative environment.

Experience

Data Science Internship | SkilledScore

March 2025 - May 2025

- Completed a data science internship focused on solving real-world problems, working on five end-to-end projects involving churn prediction, sentiment analysis, fraud detection, time series forecasting, and data pipeline development.
- Applied machine learning, NLP, and time series techniques using Python and SQL; delivered insights through dashboards and automation, enhancing both technical proficiency and business understanding.

Software Engineer | i2c Inc.

May 2023 - January 2025

- Performed data analysis and automated reporting using SQL and Python to monitor critical applications and support decision-making.
- Leveraged tools like Dynatrace, Xymon, Wazuh, and Nagios to analyze application metrics, identify anomalies, and ensure SLA compliance.

Education

BS Information Technology

October 2018 - July 2022

- Punjab University College of Information Technology (PUCIT)

Skills

- Programming Languages:** Python, SQL
- Libraries/Frameworks:** Scikit-learn, PyTorch, TensorFlow, NumPy, Pandas, Matplotlib, Seaborn, Plotly, SpaCy, Gensim, fastText
- Data Tools:** Excel, Power BI
- Concepts & Techniques:** Statistical Machine Learning, Deep Learning, Natural Language Processing (NLP), Time Series Analysis, Data Visualization, Predictive Modeling, Feature Engineering

Certifications

- Deep Learning: Beginner to Advanced - CodeBasics (May 2025)
- Smart Tips: Soft Skills for Technical Professionals - Udemy (February 2025)
- The Data Science Course: Complete Data Science Bootcamp 2024 - Udemy (November 2024)

Projects

Customer Churn Prediction | [Link](#)

Solo Project / Python, Pandas, Scikit-learn / Dataset: 7k+ rows

- Built a classification model to predict telecom customer churn using a real-world dataset, achieving ~80% accuracy on the test set.
- Performed end-to-end data preprocessing, including handling missing values, encoding categorical features (Label Encoding), and normalizing numerical columns.
- Trained and compared multiple algorithms—Logistic Regression, Decision Tree, Random Forest—and selected the best-performing model based on cross-validated F1-score.
- Used evaluation metrics such as Precision, Recall, Accuracy, F1-score, and Confusion Matrix to ensure balanced performance and minimize both false positives and false negatives.
- Generated insights on key churn indicators (e.g., contract type, monthly charges) to support actionable business decisions.

Cost of International Education | [Link](#)

Solo Project / Python, Seaborn, Plotly, Pandas, Scikit-learn / Dataset: 1k rows

- Built a regression model using an artificial neural network (ANN) to predict international education costs with ~90% test accuracy, helping uncover pricing dynamics across institutions.
- Conducted in-depth exploratory data analysis (EDA) to identify influential factors such as country, degree type, and public vs. private institutions.
- Engineered features by imputing missing values, encoding categorical variables (e.g., country, degree), and scaling numerical features to improve model learning.
- Evaluated model performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure reliable cost predictions.
- Compared ANN with baseline models (e.g., Linear Regression, Random Forest) to validate model choice and improvement.

LinkedIn Data Jobs Analysis | [Link](#)

Solo Project / Python, Seaborn, Plotly, Pandas, RegEx / Dataset: 300+ rows

- Analyzed a dataset of LinkedIn job postings to uncover hiring trends, skill demands, and regional variations in the data science job market.
- Cleaned and standardized inconsistent job titles using regular expressions, improving data quality and enabling accurate aggregation and analysis.
- Performed EDA to identify top hiring locations, most in-demand data roles, and frequently required skills using grouped statistics and pivot analysis.
- Created dynamic visualizations (bar plots, heatmaps, distributions) with Matplotlib and Seaborn to communicate insights effectively.
- Extracted and summarized technical and soft skills from job descriptions to generate a data-driven snapshot of current market expectations.

Real And Fake News Classification | [Link](#)

Solo Project / Python, Gensim, SpaCy, NLP / Dataset: 17k+ rows

- Built a binary text classification model to detect fake news using Word2Vec embeddings trained on the Google News corpus, capturing deep semantic relationships across words.
- Preprocessed raw text data through tokenization, stop-word removal, and lemmatization, ensuring clean and consistent input for embedding and model training.
- Generated 300-dimensional word vectors with Gensim's Word2Vec and averaged them to create document-level features for classification.
- Trained and tuned models such as Support Vector Machine (SVM) and Logistic Regression, improving prediction accuracy compared to bag-of-words baselines.
- Evaluated models using Accuracy, Precision, Recall, and F1-score, ensuring balanced performance for both classes.