ARFA KHALID

(639) 382- 3258 | Toronto, ON | arfakhalid09@gmail.com | linkedin.com/in/arfakhalid-ms/

DATA SCIENTIST

With 4+ years of experience in data analysis and research, I have delivered impactful results such as a 15% increase in compliance rates, 30% improvement in feature extraction efficiency, and a 20% increase in prediction accuracy. Build report automation processes and integrated Bayesian techniques, leading to a 15% boost in model performance.

SKILLS

Technical Skills: Python (Scikit-learn, pandas, NumPy, Statsmodels), R (caret, dplyr, ggplot2), SQL, Excel, statistical model building and analysis, Business analysis (Tableau, Power BI, QlikView), Spark, Machine Learning (Tensorflow, Keras), deep learning (neural networks), cloud computing , version control (Git)

Interpersonal Skills: strategic problem-solving, insight & analysis, analytical and critical thinking, communication, detail oriented, result focus, time management, work-independently and collaboratively

WORK EXPERIENCE

QTO House **Remote**

Data Scientist *June 2023 - Present*

- Generating ad-hoc reports and analyses for compliance business partners, providing them with critical insights to support their decision-making processes.
- Collaborating with key business partners to design and implement data-driven solutions that are resulting in a 15% increase in compliance rates.
- Automating scheduled reporting processes, reducing manual effort by 20% and freeing up data analysts to focus on more complex tasks.
- Developing and implementing a data pipeline using Hive and Python that integrates data from four sources and feeds a new compliance dashboard, enabling comprehensive agent performance tracking and leading to a 10% improvement in agent productivity.

University of Regina **Regina, Saskatchewan, Canada**

Graduate Research and Teaching Assistant *August 2021 – April 2023*

- Enhanced feature engineering efficiency by 30% by extracting 5x more key features from massive datasets using parallel processing and distributed computing frameworks.
- Led the development of groundbreaking models employing big data analytics and machine learning to bridge knowledge gaps in complex statistical and mathematical real-world problems, achieving a 20% increase in prediction accuracy.
- Integrated Bayesian techniques into big data models, resulting in a 15% boost in model performance, leading to enhanced predictive capabilities, improved decision-making support, and a 10% reduction in overall uncertainty.
- Instructed Laboratory Classes to undergraduate students, conducted office hours and help faculty in conducting exams and grades marking.

Pakistan Bureau of Statistics **Lahore, Punjab, Pakistan**

Data Analyst Intern *June 2019 - December 2019*

- Streamlined data collection processes, reducing collection time by 20% and improving accuracy by 15%.
- Implemented data quality checks and automated data cleansing, reducing data errors by 30%.
- Developed and maintained dashboards and reports to track key performance indicators (KPIs), driving data-driven decision-making.
- Uncovered hidden patterns in population and agricultural data, leading to targeted interventions that improved productivity and health. Collaborated with cross-functional teams to assess the impact of programs on population and agricultural trends.

EDUCATION

University of Regina, Saskatchewan, Canada **April 2023**

Master of Science in Applied Statistics

Course work: Probability, Statistical Inference, Statistical Modeling of Dependence and Extremes, Design of experiments, Bootstrap methods

University of the Punjab, Punjab, Pakistan **September 2020**

Bachelor of Science in Mathematics and Statistics

Course work: Sampling Techniques, Multivariate Techniques, Applied Econometrics, Time series Analysis, Total Quality Management, Research Methodology, Categorical Data analytics, Machine Learning, Database development and design, , computer science, artificial intelligence

PROJECTS

Climate Change Analysis: Time Series Forecasting in ML

- Led Climate Change Analysis project, utilizing Matplotlib, NumPy, Seaborn, and scikit-learn for data analysis. Built linear regression models to NASA's dataset (1951-1980) and extended insights to a broader range (1882-2014). Quantified temperature change patterns, revealing potential acceleration in climate change. Ensured model accuracy through rigorous assessment of actual vs. predicted temperature values.
- Identifying a 40% increase in global average temperature. Predicted further warming with 95% confidence, informing climate mitigation strategies.

Predicting Public Opinion with Deep Learning and Natural Language Processing

- Executed sentiment analysis on the IMDB dataset (50,000 movie reviews) using Keras for neural network modeling. Achieved an 86.97% accuracy on the test set through meticulous data preprocessing, neural network design, and optimization. Developed a versatile sentiment analysis function for custom text inputs, demonstrating the model's adaptability and generalizability.

Satellite Imagery Segmentation for Flood Impact Assessment in Libya using Computer Vision

- Led geospatial flood analysis in Libya using Maxar Open Data. Employed segment-geospatial and leafmap libraries for efficient data processing. Achieved accurate flood segmentation, converting raster masks to vectors. Visualized results on an interactive map, demonstrating insights into flood-affected areas.