ARFA KHALID

+16393823258 | Toronto,ON | arfakhalid09@gmail.com | linkedin.com/in/arfakhalid-ms/

EDUCATION

University of Regina, Saskatchewan, Canada **June 2023**
Master of Science in Statistics
Thesis : Analyzing the Effectiveness of Covid-19 Vaccines using Multinomial Logistic Regression Model

University of the Punjab, Punjab, Pakistan **September 2020**
Bachelor of Science in Mathematics and Statistics
Thesis : Qualitative Behavior of Students on Cyberloafing using Multivariate Model Analysis

PROFESSIONAL EXPERIENCE

QTO House **Remote**
Data Analyst *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, SK, Canada**
Graduate Research and Teaching Assistant *August 2021 - May 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.

SKILLS

Technical Skills : Statistical Modeling and Analysis, Geospatial Data Analysis, Python, R, Matlab, SQL , Fortran , Latex, QGIS, ArcGIS (GIS data/mapping)

Soft Skills : Communication, Problem-Solving, Probabilistic assessment, Collaboration, Attention to detail , Adaptability, Continuous Learning, Critical Thinking

PROJECTS

Flood Analysis using SAMGEO

- Utilized Samgeo Computer Vision model for deep learning to segment satellite imagery of Libya, showcasing the effects of flooding caused by Storm Daniel and the collapse of two dams near Derna. Transformed the segmented imagery into an interactive map, providing a comprehensive visual representation of the destruction caused by the event, using Maxar Open Data.

Climate Change Analysis

- Advanced data analysis and Python's Seaborn skills to analyze 30 years of NASA climate data. Identified a 40% increase in global average temperature and predicted further warming of more than 45 degrees Celsius by 2030 with 95% confidence.

Proximity Analysis for EPA Environmental Monitoring

- In this project for the Environmental Protection Agency uses GeoPandas and Folium to analyze toxic chemical releases in Philadelphia. Key findings include a mean distance of 33,516 feet between release incidents and monitoring stations, visualized with a 2-mile buffer around each station on a map. Efficient testing reveals proximity of toxic releases to monitoring stations.