Muhammad Arslan Siddiqui | MS – Data Science
Karachi, Pakistan
0092-331-3113789 | muhammadarslan.se@gmail.com
LinkedIn: https://www.linkedin.com/in/muhammadarslansiddiqui/
Portfolio: https://muhammadarslanidrees.github.io/

Professional Profile:

Highly analytical, technically astute and Professional Data Scientist with demonstrated history of over 4+ years' experience of successful development and deployment of predictive models on large scale in financial industry. Working in capacity of Manager, Data Science, I adept in the end to end delivery of value-added use-cases and operational improvements based on effective requirements identification / analysis through effective stakeholder engagement. Communicative, diligent, and thoroughly dedicated in all pursuits. Broadly, I've expertise in Data Warehousing, Cloud Architecture and Artificial Intelligence.

Technical Expertise:

Python | R | Java | .Net | SQL | Oracle | Microsoft SQL Server | MySQL | TensorFlow | PyTorch | Keras | Scikit | Spark | Hadoop (MR) | Impala | Hive | YARN | Qlikview | Tableau | Power BI | Rapid Miner | Docker | Linux | Cloudera | MS Project | Big Data Processing | GCP | Data Mining | Statistical Analysis | Data Warehousing | Cloud Architecture | Artificial Intelligence

Education & Qualifications:

2017 – 2019: FAST – National University of Computer & Emerging Sciences, Karachi, Pakistan | MS Data Science | CGPA: 3.51

MS Thesis: Addressing Anomaly Detection in High Volume Setting – https://github.com/muhammadarslanidrees/AML_THESIS

The research aims to solve the biggest challenge in fintech industry i.e. Money Laundering Detection, through advanced deep learning techniques. Using bank's real data, a data driven system was developed which deals with class imbalance using Deep Generative Models and uses proposed technique "Cramer Rao Lower Bound Optimization" for handling high volume and high velocity transactional data.
Technology used: Python, SQL, Keras, Pytorch, TensorFlow, Jupyter, JSON, SQLAlchemy.

Publications – https://www.researchgate.net/profile/Muhammad_Siddiqui18

- Sustaining learning under density shift: a Big Data volume to velocity reduction
- Deep generative models to counter class imbalance: a guided model selection strategy

Project: Sentiment Analysis on Tweets – The main components are tweets extraction, text analysis, hashtag analysis and developing a classification model for tweets sentiment.
Technology used: Big Query, Compute Engine, NLP API, Google Dataflow & Data Studio, Twitter API, JSON.

Project: Data Warehouse of Retail Company – The main components are analyzing the datasets, ETL, dimensional model with dimensions and measures, ingest data in dw tables and write analysis queries (data mining).
Technology used: MS SQL Server, MS Visual Studio, SQL, MS Excel, Pentaho Data Integration.

2012 – 2015: SIBA – Sukkur Institute of Business Administration, Sukkur, Pakistan | BS Software Engineering

Achievements:

- Awarded Gold Medal on securing 1st Position in BS degree.
- Awarded National ICT R&D Scholarship (a 4-year need cum merit based full scholarship)
- Appeared in Deans list for outstanding performance in academics
- Awarded Bait-ul-Hikma award for conducting workshop as Trainer at Hamdard University Karachi

Employment Overview:

Jan 2019 – Present: HBL – Habib Bank Limited, Karachi, Pakistan | Manager Advanced Analytics
HBL is a Karachi based multinational bank. Working as Senior Data Scientist, my responsibilities include:

- Manage and lead data science team on critical projects (technical)
- Formulate analytics opportunities by engaging with other bank units
- Work as a liaison between technical team and management (convert ML jargons into explainable insights)

Achievements:

- Developed data science team
- Setup data science Infrastructure & workflows
- Operationalize data science use-cases
- Developed Non-traditional credit scoring model
- Optimized thresholds for Money laundering detection

Feb 2018 – Dec 2018: HBL – Habib Bank Limited, Karachi, Pakistan | Data Mining Analyst
As Data Scientist / Engineer, my responsibilities include:

- Research on Machine learning techniques, data exploration, ETL, Feature engineering, model building & deploying use-cases on Data Lake.

- Implement data science pipeline using O.S.E.M.N framework.
- Develop effective analytics in collaboration with other units and present insights to the business.

Project: Non-traditional Credit Scoring Model – Scoring model was developed using logit and weight of evidence method was used for transformation of non-traditional features. The scorecard was also validated using population and characteristic analysis.

Project: Segmentation – Anti-Money Laundering – Applied clustering to segment customers based on their transaction behavior and explained segments using rules.

Project: Thresholds Optimization – Anti-Money Laundering – Calibrated Transaction Monitoring System Scenarios using Statistical techniques and analyzed the impact using ATL/BTL testing in order to improve a financial institution's capability to detect suspicious activity and reduce false positive alerts.

Project: Alert Risk Scoring – Anti-Money Laundering – In order to increase the operational efficiency of AML/Compliance analysts, prioritized money laundering alerts by developing a Scoring Model using ML algorithm which assigns a risk to each alert and categorize among high, medium and low risk.

Project: Cross Sell Bank Products – In this project, Tree based Machine Learning algorithm was used to identify potential customers likely to buy personal loan or credit card in order to enhance the pool of customers.

Project: Screening Customer's Portfolio with Fuzzy Matching – In this project, Levenshtein distance was used as fuzzy matching algorithm for screening customer's portfolio against FATF entities to mitigate the risk of terrorist financing.

Project: Anomaly Detection in Agent Transacting Pattern – Developed model to identify and report those branchless banking customers who depict anomalous transaction behavior. Experiments using different anomaly detection techniques were carried out but Isolation Forest proved to be the best as it is easy to interpret.

Apr 2016 – Feb 2018: CDCPL – Central Depository Company Pakistan Limited, Karachi, Pakistan | Software Engineer & BI Expert
At CDC, I worked on Java technologies and contributed in all phases of the development life-cycle. Also, developed an end to end BI Application on Qlikview which includes writing data modelling scripts, designing and deploying dashboards.

Project: Reporting Automation – Automated the reporting process by developing a BI Application which includes incremental data ingestion, data modelling, transformation and designing interactive dashboards and deploying on Qlikview Server
Technology used: Oracle 12c, Qlikview, SQL.

Project: Registrar & Transfer Agent (RTA) – This is a book entry service providing solution of Central Depository Company to its clients. I contributed in developing different reports and enhancing system capabilities using Java.
Technology used: JSF2.2, PrimeFaces6, BIRT, Web Logic Server, Oracle 12c.

Project: Central Depository System (CDS) – CDS is electronic book-entry system used to record and maintain securities and to register the transfer of securities, Using Java as backend and oracle database, I contributed in extending core system with new functionalities and reports.
Technology used: Enterprise JavaBeans, Web Logic Server, JReports, Oracle 12c.

Jun 2014 – Jul 2014: Microsoft Innovation Center, Lahore, Pakistan | Intern
- Worked on Microsoft technologies including .Net, Windows app development and Unity 3d.

Certifications:

- End to End Machine Learning with TensorFlow on Google Cloud Platform | **Coursera**
- Specialization: TensorFlow in Practice | **Coursera**
- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning | **Coursera**
- Convolutional Neural Networks in TensorFlow | **Coursera**
- Natural Language Processing in TensorFlow | **Coursera**
- Sequences, Time Series and Prediction | **Coursera**
- Deploy Models with TensorFlow Serving and Flask | **Coursera**
- Analyzing Business Requirements for Data Science | **PluralSight**
- Building Web Applications in R with Shiny | **DataCamp**
- Python for Data Science | **DataCamp**
- Data Visualization | **DataCamp**
- Machine Learning | **DataCamp**
- R | **DataCamp**

Blogs:

- Towards Data Science: https://towardsdatascience.com/@muhammadarslan.se
- Microsoft TechNet: https://social.technet.microsoft.com/profile/muhammad%20arslan%20siddiqui/

Languages: English | Urdu