Arushi Sharma

(914) 484-8552 , Boston, MA

Website | Github | LinkedIn | sharma.aru@northeastern.edu

SUMMARY

A passionate data science professional, with over 5 years of industry experience working in financial, telecom and technology domains, with an aim to drive optimal decision making for real-world business problems

EDUCATION

- **Masters (MS in Data Science)**; Northeastern University, Boston, MA   *Expected − Dec,* 2021
  Relevant Courses: Algorithms, Linear Algebra and Probability, Information Retrieval, Machine Learning
- **Bachelors (B.Tech in Computer Science)**; Rajasthan Technical University, India, **Honours**   *May,* 2012
  Relevant Courses: Data mining and Warehousing, Distributed Systems, Statistics and Probability Theory

TECHNICAL SKILLS

- **Programming:** Python, Shell, SQL, R, Terraform, Numpy, Pandas, Scikit-Learn, SciPy
- **Technologies:** Hadoop, Spark, Kafka, Hive, Cloud (AWS), Docker, Elastic Search, PyTorch, Tensorflow
- **Machine Learning:** Regression (Linear, Logistic), Regularization, Decision Trees, Boosting, Bagging (Random Forest), SVM, Naive Bayes, PCA, CNN, RNN (LSTM), NLP

ACADEMIC PROJECTS

Analysis of COVID-19 spread and Fatality prediction *(R, SIR model, SVM, Random Forest, SVM One-class)*
- Predicted patient's fatality with 85% accuracy and performed data analysis, wrangling and visualization on COVID-19 cases across the world

Document Ranking using IR models *(Python, spacy, NLTK, Elastic Search, NLP, Machine Learning modeling)*
- Preprocessed, indexed and ranked documents against user queries and evaluated performance against vector-space, language and machine learning models, resulting in the improvement of average precision by 30%

Names Classification using character-level LSTM model *(PyTorch, LSTM, Scikit-learn, Numpy, Matplotlib)*
- Trained a character level LSTM model in Pytorch to predict the language of the names with 90% accuracy

Image Classification using CNN *(Python, Numpy, PyTorch, Tensorboard, Logistic Regression, CNN)*
- Created a classification model using Convolutional Neural Networks to classify objects with 70% accuracy

PROFESSIONAL EXPERIENCE

Bridgei2i Analytics Solutions, India - Senior Analytics Consultant   *Apr* 2017 - *May* 2019
Bridge Funnel (*Python, Docker, Terraform, AWS : S3, EC2, Elastic Beanstalk, ElastiCache, Lambda, Jenkins*)
- Optimized the creation of AI sales product cloud infrastructure, reducing the setup time by 60% using automation
- Containerized algorithms and configured Jenkins for continuous builds, reducing man power by 50 hours/week

Image Analytics (*Python, Kafka, Spark Streaming, Cassandra, Tensorflow, Docker, AWS*)
- Implemented an end to end object detection and classification usecase, ingesting live feed from webcam using Kafka
- Integrated algorithms with AWS, increasing scalability and cost efficiency while reducing the execution time by 30%

Anomaly Detection (*Python, PySpark, Spark Streaming, Cassandra, Kafka*)
- Designed and created data pipelines for anomaly detection by collecting data from various sensors
- Performed data cleaning and manipulation on sensor data, using it for real time monitoring of the anomalies

Lead Engine (Intent Marketing) (*Pyspark, AWS, DataBricks*)
- Optimized the code in PySpark to run on Spark 2.x version and improved the performance of Lead Engine, lowering the real-time execution by 75%
- Designed a level control framework in the application, controlling multiple execution points from a single program

Accenture, India - Application Development Analyst   *Feb* 2016 - *Apr* 2017
Data Architecture and Warehousing (*Hortonworks, Diyotta, SQL, HDFS, Hive, Hue, Sqoop, Flume*)
- Generated business insights by creating data pipeline and consuming historical and incremental data from multiple sources to Hadoop data lake
- Achieved reduction in the operation cost by 35% by building automated pipelines and hive scripts using Diyotta

IBM, India - Software Developer   *Jan* 2013 - *Feb* 2016
DCR Analysis (*Cloudera, Hadoop, Hive, Sqoop, Oozie, Pig, SQL*)
- Developed hive scripts to generate insights from call center data, improving dropped call rate by 25%
- Improved query performance by 40% by using hive partitioning and bucketing after migrating data to HDFS
- Researched and implemented language translation module in Java, increasing revenue and client satisfaction
- Collaborated with key business stakeholders to understand requirements and led ideation sessions in the team

ACHIEVEMENTS

- **Certificate of Recognition, Bi2i** for mentoring analytics professionals in Big data, Pyspark and Cloud   2018
- **Managers Choice Award, IBM** for excelling in the practice of "Put the Client First"   2015
- **Orion Award Winner, IBM** for exemplary performance and exceptional dedication to IBM   2014
- **Outstanding Contributor Award, IBM** for excellent performance for the year 2013-2014   2014