Abdullah Biswas

Data Scientist

San Francisco Bay Area, CA - Email me on Indeed: <u>indeed.com/r/Abdullah-Biswas/09837f6421880e6e</u>

Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

Indiana University Bloomington - August 2015 to May 2017

Kaggle: House Prices: Advanced Regression Techniques (scikit-learn)
Built Machine Learning Models to predict House Prices. Preprocessed the dataset, constructed new features, selected features and evaluated the models - RandomForest, GradientBoosting, Ridge, Lasso and XGBoost.
• Ranked in top 10% in the public leader board.

Machine Learning Projects (Python, NumPy, Pandas)
Developed Machine Learning Models from scratch - k-Means Clustering, Gaussian Naïve Bayes, Logistic Regression, Neural Network Classifiers, Linear Regression, and Kernel Regressions with Regularizers.

Recommendation System Project (Python, Spark)
• Developed a Movie Recommendation System, using Apache Spark MLlib, HDFS, Python and Ansible on FutureSystems, a high-performance cluster computing infrastructure.
• Handled MovieLens' dataset containing more than one million movie ratings.

Hadoop Projects (Java)
• Implemented parallel version of PageRank algorithm using Java and Hadoop MapReduce Framework.
• Developed index table using HBase and built a Search Engine using the index table and the PageRank.

Data Visualization Project (Python, Tableau, scikit-learn)
• Created visualization of Temporal Sentiment Analysis of Ambition, the Film for a Client (Leiden Observatory).
• Retrieved YouTube comments using Google Data APIs, preprocessed using Python, analyzed Sentiment using MultinomialNB classifier and visualized the temporal sentiment by Tableau.

Technical Lead

Aricent Technologies - June 2012 to July 2015

• Ensured high quality of Cisco's Routers and Aricent's Switching Solution by Testing and Automation.
• Led a team of software engineers and doubled its productivity based on bug detection and verification.

Cisco Systems Software Engineer
• Improved quality of Cisco's MDS Storage Networking Switch by software testing and Automation.

EDUCATION

Master of Science in Data Science

Indiana University Bloomington
August 2015 to May 2017

Bachelor of Engineering in Computer Science
National Institute of Technology

SKILLS

MACHINE LEARNING (2 years), APACHE HADOOP MAPREDUCE (1 year), Python (2 years), Java (2 years), scikit-learn (2 years), Pandas (2 years), NumPy (2 years), R (1 year), Spark (1 year), SQL (1 year)

ADDITIONAL INFORMATION

• Data Scientist/Software Engineer with hands-on experience in development of accurate Machine Learning Models, Deep Learning Tools, Data Analysis, Visualization and in Data Science Tools.
• Software Development expertise in Java, Python, and C. Experience in Software Testing and Automation.

TECHNICAL SKILLS
• Languages: Java, Python, C, R, NumPy, SciPy, Pandas, Matplotlib, SQL.
• Framework & Tools: Spark, Hadoop, MapReduce, Guess, Tableau, Git, Ansible, Eclipse, XGBoost.
• Machine Learning: Classification, Regression, Clustering, Feature Engineering, MLlib, scikit-learn, TensorFlow, Markov Model, HMM, Probability Theory, Statistical Inference, Hypothesis Testing.