Indraneil Bardhan

Summary

Data science generalist with experience in developing leading edge machine learning models on a wide range of big data platforms and unlocking potential hidden in large scale data systems. Eager to bring expertise and fresh ideas to growing organization in challenging new role.

Skills

Focus Areas: Statistics, Regression, Topic Modeling, Time Series Analysis, Natural Language Processing (NLP)
Programming: Python, scikit-learn, h2o, pandas, numpy, SciPy, geopandas, SHAP, LIME, matplotlib, seaborn, R, tidyverse, ggplot2, plotly, mlr3, SQL
Frameworks: TensorFlow, Apache Spark, Hive/Impala, Amazon Web Services (AWS), Google Cloud Platform (GCP)

Professional Experience

Data Scientist - r4 Technologies *April 2017 - Present*

- Prototyped and implemented highly scalable, end-to-end volumetric forecasting pipelines with **SciPy**, **PySpark** and **statsmodels**, for XEM; R4's proprietary AI product.
- Developed and evaluated demand forecasting methodologies using **ARIMA**, **Holt-Winters** and **XGBoost** on **PySpark** to help develop quarterly roadmaps to drive asset placement for a lumber manufacturer, leading to an overall lift of $3M.
- Prototyped a scalable **topic modeling** system on **PySpark** using **NLTK** and **LDA** on **gensim**, for an executive talent acquisition firm, improving time efficiency by 50%.
- Implemented survival analysis using **lifelines** and **Scikit-learn** for a manufacturing company, reducing unplanned stoppages by 35%.
- Designed and deployed a **product recommendations** workflow to feed email promotions for an online sports retailer using **K-means** and **Collaborative Filtering**, doubling the conversion rate.
- Scaled a location intelligence model for a credit card provider, using **Rtree**, **shapely** and **geopandas** libraries, increasing coverage area by 80%.

Research Associate - University of Pittsburgh *September - December 2016*

- Curated and visualized self-reported public health data using **dplyr** and **ggplot2** to support research findings for publications.
- Designed and built statistical models using **pandas**, **numpy** and **scipy** to evaluate the relationships between food insecurity, depression and diabetes.

Statistical Programming and Analysis Intern - Genentech Inc. *June - August 2016*

- Reduced the model computation speed of a classification model 8% by using PySpark on a custom experimental hadoop ecosystem, while keeping the error rate constant.
- Developed A/B tests to compare custom experimental hadoop ecosystem with the legacy analysis frameworks.

Education

M. S. Information Science - University of Pittsburgh *2015 - 2016*

- *Concentration* : **Big Data Analytics** *Coursework* : Cloud Computing, Machine Learning, Advanced Databases, Information Visualization, Statistical Learning and Algorithm Design
- GPA: 3.458

B. Engg. Computer Engineering - University of Pune *2011 - 2015*

- *Coursework* : Software Development, Object-Oriented Design, Artificial Intelligence, Engineering Math
- GPA: First Class

indraneil.bardhan@gmail.com *203-900-4061* *412-961-5794* *Stamford, CT*