ANKIT BELADIYA

Data Scientist

Toronto, ON

514-770-3903 | ankitbeladiya.ca@gmail.com | https://github.com/ankitbeladiya | https://www.linkedin.com/in/ankit-beladiya/

Soft Skills

- Data Scientist with industrial experience in building and deploying AI based financial product in production
- Problem-solver, Organized, Detail-oriented, Autonomous, Quick learner

Technical Skills

Machine Learning	Classification, Regression, Clustering, Feature Engineering, Hyper-parameter Optimization
Python	Pandas, Scikit-Learn, SciPy, NumPy, OpenCV, Matplotlib, Seaborn, Tensorflow, Keras, Joblib, SpaCy, Dask
Spark	PySpark, MLlib, SparkSQL
Cloud	AWS
DB	PostgreSQL, MongoDB
DevOps	Docker, Docker Compose, Kubernetes
IDE	PyCharm, Jupyter, Zeppelin
OS	Linux, MacOS
Tools	Jira, Confluence, Github, Lucidchart

Education

University of Windsor — *Windsor, ON*

Master of Engineering - Electrical [Data Science applied to Electrical] — *May 2017 – Aug 2018*

Experience

Stradigi Ai — *Montreal, QC*

Data Scientist — *Jan 2019 - March 2020*

- Problem formalization based on client's requirements
- Data collection, cleaning and storing in database
- Data exploration, visualization and reporting
- Build data science pipelines for time series forecasting, NLU and image processing
- Hyper parameterization of classical machine learning models and neural networks
- Data parallelization using Dask
- Write reports being communicated to the client
- Build and deploy models on AWS EC2 using Docker
- In a team of 3 data scientists, build Automated Trading Platform using supervised and unsupervised ML algorithms
- In a team of 2 data scientists, build Image Segmentation Pipeline using deep learning for a client in Health Care
- Worked using Agile methodology
- Followed software development best practices
- Provided training of machine learning and big data concepts

Projects

Missing Data Imputation — *Aug 2017*

- Implemented fuzzy clustering-based algorithm using Scikit Learn
- Used Elbow method to find number of clusters in data
- Imputed missing data using Expectation Maximization (EM) method
- Compared algorithm's accuracy and performance against K-means and Gaussian Mixture Model (GMM) clustering
- Developed python API for easy implementation of the algorithm

Recommendation System with Big Data — *Apr 2018*

- Uploaded user data from local file system to Hadoop file system (HDFS)
- Read the data from the HDFS using Apache Spark
- Data modelling and preprocessing using PySpark and Spark SQL library
- Applied the Collaborative Filtering Technique to recommend new products using Spark MLlib
- Stored results in Hive Database

Certificates

Coursera	Applied Data Science with Python	*Sep 2018*
Coursera	Machine Learning with Big Data	*Sep 2018*