SATISH REDDY CHIRRA

(857)-237-7974 • Boston, MA – 02120 • chirra.s@husky.neu.edu
LinkedIn: www.linkedin.com/in/satish-chirra • GitHub: github.com/chirras • Website: chirras.github.io/

Available for Full-time positions starting from September 2019

EDUCATION

Northeastern University, Boston, MA
Candidate for a Master of Science in ***Data Science*** **GPA: 3.4/4** Jan. 2017 - Present
Related Courses: Machine Learning, Data Mining, Algorithms, Cloud Computing, Data Visualization, Information Retrieval, Data Management & Processing, Database Management System

Jawaharlal Nehru Technological University, Telangana, India
Bachelor of Technology in Electronics and Communication Aug. 2009 - May 2013

TECHNICAL KNOWLEDGE

Key Strengths:	**Feature Engineering, Time Series Forecasting, Natural Language Processing, Data Mining, Deep Learning, Data Visualization, Statistical Modelling, ETL, Recommendation System**
Languages & ML Tools:	R, Python (Scikit-learn, Numpy, Pandas, PySpark), Tensor Flow, Keras, Weka, Flask
ML/AI Platforms:	Data Robot, H2o Driverless AI, AWS SageMaker, BigML, Azure AML
Cloud Platforms:	AWS, Google Cloud Platform, Azure, Heroku, Elasticsearch, Kubernetes, Docker
Big Data Technologies:	Spark (PySpark, MLIB, Spark SQL), Hadoop
Visualization & Interpretability:	D3, Tableau, matplotlib, Plotly, ggplot, Shiny; LIME, SHAP
ETL Tools:	SAP BODS, Informatica, SAP BW, HVR
Database:	SQL (Oracle, Vector-Wise, SQL Server, Teradata), NoSQL (MongoDB), Google Big Query
Functional Skills:	Agile Methodology, Waterfall Model
Others:	Linux, Java Script, HTML, Microsoft Excel, Access, JIRA, Service-now, Putty

PROFESSIONAL EXPERIENCE (~ 4 Years)

Fidelity Investments May 2018 - Dec. 2018

Data Scientist Coop – AI Team, Boston, MA

Stock Hard-to-Borrow Prediction (*Multi Time-series Classification*)

- Engineered time related features like Lags, Moving Averages, Difference columns and TF-IDF Vectors on Text data.
- Built ML predictive models (XGBoost, LSTM) in python and used Backtesting for cross-validation and Walk forward validation techniques which resulted in 1-10 days ahead prediction and better planning of inventory for brokers.
- Interpreted model predictions using LIME to explain features contributing positively for the prediction to brokers.

Financial Wellness Score (FWS) - eMoney

- Implemented Financial Wellness Score for eMoney clients identifying critical categories that effect financial wellness, which helps advisors to manage clients effectively and recommend Next Best Action (NBA).
- Developed interactive visualizations and animations using Plotly on various financial wellness factors.

Capgemini Nov. 2013 – Sept. 2016

Senior Software Engineer – Business Intelligence Team, **GE**, Bangalore, India

- Optimized ETL pipeline to improve the performance of loading financial data into data warehouse tables (Vector Wise) using Informatica, SAP BODS from multiple sources.
- Developed complex SQL queries to transform the data by applying the business rules before loading to DWH.
- Built Tableau dashboards to visualize financial reports & managed tableau server (user management, scheduling).
- Participated in sprint planning, daily scrums, testing, retrospectives and sprint reviews.
- Received 'PAT ON BACK' award for implementing ETL optimization techniques in different projects.

ACADEMIC PROJECTS

Image Categorization (*Stanford Dog Dataset*) Sept. 2017 – Dec. 2017

- Designed and Implemented end to end Convolutional Neural Network (CNN) for 120 different dog breeds.
- Fine-tuned a pre-trained model (InceptionV3) on 120 classes and achieved an accuracy of 99.16 on test data.

Data Mining Approach to Identify Air Quality profiles in USA Sept. 2017 – Dec. 2017

- Extracted the data from EPA Air quality data via Google cloud platform using SQL.
- Implemented K-Means to cluster different pollutants across the states in USA both season and site wise. Also, used K-Median, PAM, Hierarchal, DB Scan to compare the results.

Detection of Fake News Posts on Facebook Jan. 2017 **–** April. 2017

- Crawled Facebook news post comments using API and tokenized into a 300-dimensional vector space.
- Trained the vectors on the corpus of comments, also used pre-trained vectors on 3 billion words from Google.
- Implemented machine learning algorithms like Random Forest and achieved an accuracy above 90%.