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, Docker & Container Technologies
Big Data Technologies:	Spark(PySpark, MLIB, Spark SQL), Hadoop
Visualization & Interpretability:	D3, Tableau, matplotlib, Plotly, ggplot, Shiny; LIME, SHAP, Global Surrogate
ETL Tools:	SAP BODS, Informatica, SAP BW, HVR
Database:	SQL (Oracle, Vector-Wise, SQL Server, Teradata), NoSQL (MongoDB), Google Big Query
Others:	Java Script, HTML, ETL Techniques, Excel, Microsoft Office and, JIRA, Service-now, Putty

PROFESSIONAL EXPERIENCE

Fidelity Investments   May 2018 - Dec. 2018

Data Science Co-op – AI Team, Boston, MA

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

- 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 Back testing 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.

Technologies Used: Python | Ensemble Algorithms | RNN (LSTM) | Feature Engineering | LIME | SMOTE | Plotly

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.

Technologies Used: Python | Pandas | Plotly | Financial Modeling | Data Analytics | EDA

Cryptocurrency Analysis

- Conducted thorough exploratory data analysis to find the patterns between cryptocurrency exchanges.
- Built a process flow to analyze market behavior using data from different exchanges and find the best price of bitcoin at a given time.

Technologies Used: Python | EDA | Pandas | Plotly | Market Analysis

Proof of Concept – ML Platforms

- Conducted POC across different Auto-ML platforms like Data Robot, H2o Driverless AI, BigML and AWS Sagemaker.

Capgemini   Nov. 2013 – Sept. 2016

Senior Software Engineer – Business Intelligence Team, Bangalore, India

Client – General Electric Power & Water   Feb. 2015 – Sept. 2016

- Optimized ETL pipeline to improve the performance of loading financial data into data warehouse tables (Vector Wise) within 4 hours using SAP BODS from multiple sources, on which various other jobs depends.
- 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).
- Conducted UAT with the business teams to ensure the system was aligned with the vision of the business teams.

- Participated in sprint planning, daily scrums, testing, retrospectives and sprint reviews.
- Received 'PAT ON BACK' award for implementing optimization techniques in different projects.

Technologies Used: ETL | Advanced SQL | Tableau | HVR | Vector Wise | Tableau Admin | Unix | Agile | JIRA

Software Engineer
Client – General Electric Capital America, Bangalore, India   May 2014 – Jan. 2015

- Optimized ETL Job which takes more than 90mins to load History data (134Million records) to 4mins using Parallel Partitions at session level of Informatica and also incremental (Changed data) data to 1min which loads to both Oracle and Teradata databases from a flat file.
- Performed data validation by developing and executing test plans and supporting user acceptance testing.
- Improved the database performance by indexing appropriate columns in various tables.
- Optimized the database(oracle) by partitioning at table level for performance improvement.

Technologies Used: ETL | Oracle | SQL | Informatica | Indexing |UAT | Agile | Unix

---

ACADEMIC PROJECTS

Forecasting Retail Sales Data   Jan. 2019 – April. 2019

- Built models using both the time series (ARIMA), supervised machine learning techniques (XGBoost) and Deep Learning (LSTM) leveraging different Feature Engineering techniques.
- Developed visualizations for trends in sales and predictions across different locations using D3.js V4.
- Designed and Developed an application 'Sales Analyzer' and deployed it using Heroku.

Technologies Used: D3.js V4 |Python | Tableau | Keras | JavaScript | HTML | CSS

Cloud Computing   Jan. 2018 – April. 2018

- Developed a pipeline which allows researchers/users to run the Data Science projects using Google Cloud Engine.
- Automated the process of creating the virtual machine, installing the required environment and packages and running the project on cloud using Docker Images and Containers.
- Collaborated in developing a front-end application where users can register their projects with the required details and keep a repository of their work along with the research papers.

Technologies Used: Google Cloud Platform (GCE) | AWS | Azure | Python | JavaScript | Docker | Containers

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.

Technologies Used: R | EDA | Keras | CNN | Inception V3 | ggplot

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.

Technologies Used: Python | Google Big Query| SQL | Clustering Algorithms | Matplotlib | EDA

Information Retrieval   May 2017 – Aug. 2017

- Developed a 'Web Scraper' that uses BFS algorithm on the seed URLs by prioritizing the links to be crawled with the highest number of in-links and crawled above 30000 URLs in Python and indexed using Elasticsearch.
- Implemented a Spam Classifier using Machine Learning Algorithm (Decision Trees) and Elasticsearch.

Technologies Used: Python | Elasticsearch | Kibana | Web Crawling | Indexing | Decision Trees | DFS | BFS

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, Nearest Shrunken Centroid and achieved an accuracy above 90%.

Technologies Used: R | Word2Vec | Random Forest | Bayesian Inference | Google word vectors | NLP

Sentiment Analysis on HealthCare Tweets   Jan. 2017 – April. 2017

- Crawled 15000 tweets related to American Healthcare using Twitter API in R.
- Processed the corpus of words using N-gram (Uni and Bi-grams) models and extracted the sentiment of the tweets that ranging from highly Positive to highly Negative and Visualized the sentiment.

Technologies Used: R | NLP | Sentiment Analysis | Text Analysis | Word Cloud | ggplot