

# Shreya ayyangar

## Data Scientist

Lindstrom, MN

-Email me on Indeed: <http://www.indeed.com/r/Shreya-ayyengar/0e067cd8d6298fb1>

- Passionate Data Engineer with 6+ years of analytics experience in BFSI domain with excellent team building and management skills.
- Worked as Artificial Intelligence Intern. Won 1st place in UNICEF- IAS Data Analytics Hackathon, designed interactive dashboard showcasing 5 important child development indicator factors across country.
- Skilled in Machine Learning, Natural Language Processing, Data Mining, Data Modeling, Data Visualizing, Data Management, Relational and Non-relational Database. Business Analytics Graduate (MS) in data science specialization, with a demonstrated history of delivering large, complex projects.
- I have a passion for technology and the positive impact it creates in business transformation and process improvement.
- Strong analytical skills with the ability to collect, organize, analyze, and disseminate large datasets with attention to detail and accuracy
- Experienced in designing, coding, debugging, reporting, analyzing data utilizing Python worked on IDE like PyCharm, Visual Studio Code and Jupyter Notebook
- Expertise in Python (2.x/3.x) programming with multiple packages including NumPy, Pandas, Matplotlib, SciPy, Seaborn and Scikit-learn.
- Demonstrated skills to implement and develop machine learning or deep learning models in scikit-learn or TensorFlow such as decision trees, linear/logistic regression, naive Bayes, SVM, PCA, Neural Network
- Good Knowledge in Big Data with pyspark, Spark API; Spark SQL, Spark Data frame, UDF
- Hands-on experience on various function, transformation and actions performed on SparkRDDs
- Experienced in importing data onto Hadoop Distributed File System using MapReduce and performed analysis using Hive and reported insights using Tableau 2020.1
- Worked with different File Formats like excel, CSV, Parquet and Avro for Hive querying and processing
- Hands on experience on visualization tools and performing Exploratory Data Analysis (EDA) using Tableau 2020.1, matplotlib and seaborn in Python and GGPlot in R
- Generated Tableau visualizations and dashboards using tableau desktop. Generated Dashboards with Quick filters, parameters and sets to handle views more efficiently
- Combined visualizations into Interactive Tableau Dashboards and published them to the web portal such as Tableau public and Embedded Tableau interactive dashboard into webpage.
- Extensively used advance chart visualizations in Tableau like Dual Axis, Box Plots, Bullet Graphs, Tree maps, Bubble Charts, Waterfall charts, funnel charts etc., to assist business users in solving complex problems.
- Experience working in a Test-Driven Development environment and SDLC such as Agile (Scrum)
- Working on AWS cloud practitioner certification.
- Excellent communication skills, self-motivated with a high degree of attention, a good Team Player, can work independently committed to work and have sound leadership qualities along with good project managing skills

Willing to relocate: Anywhere

Authorized to work in the US for any employer

## Work Experience

---

Next Capital Tech LLC - Lindstrom, MN

January 2020 to Present

The project is about fetching cryptocurrency data from Binance exchange site and developing a neural Network model which can predict whether a user should sell, buy and do nothing. With this, we scraped article related to cryptocurrency from various social media platform using Beautiful soup Toolkit and analyzed sentiment around each cryptocurrency.

Responsibilities:

- Extracted data from a cryptocurrency exchange site binance, preprocessed it and design model which can predict with high accuracy whether a person should buy/sell or do nothing.
- Trained models using Google cloud platform by creating VM instance and setting up configuration such as NVIDIA, GPU.
- Performed Real time data analysis on test data set by validating it across created model which gave nearly 99 % accuracy
- Performed Web scraping using Beautiful Soup using Twitter API to extract Data related to various Cryptocurrency
- Performed sentiment Analysis on scraped data to find out Polarity, Subjectivity, and Emotion related to each cryptocurrency and overall market sentiment
- Used python Libraries such as vader, SentimentIntensity Analyser, TextBlob, NRCLEx
- Experimented with various NLP techniques (stemming, lemmatization, n-grams, TFIDF, etc.), machine learning models (Logistic Regression, SVM, Navie Bayes, etc), and deep learning models (Word2Vec, CNN, RNN); found the best model based on k-Fold cross validation scores.

Environment: Pycharm, Beautiful Soup, API, Google cloud Platform, Google colab, NLP.

### **Big Data Hadoop Developer**

Ericsson USA - Plano, TX

June 2018 to December 2019

To create information pool for Ericsson employees and provide them with hands on information of various Categories, issues pertaining to reaching out to relevant teams, competitor analysis and price analysis etc.

Responsibilities:

- Enhanced and Designed AI assistant using NLP techniques to create smooth interactions across various stakeholders and suppliers, develop applicable python actions to return intended information
- Used Pycharm to create and execute shell script as well as to write python classes
- Created database to connect Ericsson internal site to AI Assistant
- Connected Ericsson FAQ Page with AI Assistant which will result in reduced time to gather relevant information
- Created various intents, stories and Python scripts based on the RASA Natural Language Processing to train the model so that it can generate message with approximately 80% Accuracy
- Created multiples interactive dashboards with different KPI's to plug and play and see the data at different hierarchies
- Used all the show-me Visualizations from Tableau and created interactive worksheets, dashboards and story point to create stories

Environment: Tableau, Window server, SQL Server, Pycharm, Rasa NLU, Rasa Core, Shell Script

## **Data Engineer**

University of Texas at Dallas - Dallas, TX

January 2017 to May 2018

### **Responsibilities:**

- Performed Data Acquisition, Data preprocessing and transformation, Data Validation, Predictive Modeling and Data Visualization.
  - Conducted statistical analysis of 4K+ data and created visualization using Python (Matplotlib) to identify correlation and outliers
  - Designed classification models to predict coronary heart disease in highly imbalanced dataset. Used ensemble techniques (Bagging and Boosting method) to further improve accuracy
  - Used Evaluation matrix such as AUC, F1 and recall score
  - multiple classification models (KNN, SVM, Kernel SVM, Decision Tree, Random Forest, Logistic Regression) using Python and run grid search to find the best hyperparameters to predict the occurrence of heart disease with 94% accuracy.
  - Compared various model PCA, Random Forest, SVM, KNN using ROC curve to identify best model by AUC value
  - Developed and compared multiple regression models (Linear, Ridge, lasso, KNN, SVM, Kernel SVM, Decision Tree) using Python and run grid search to find the best hyperparameters to predict the life expectancy using RMSE value as evaluation matrix.
  - Applied Principal Component Analysis on the same dataset for dimension reduction of the feature space to make the model run faster with the same accuracy. Ran deep learning models, using optimal epochs and batch-size for classification
  - Saved Model in Pickle file and deployed same using Flask API on Heroku.
  - Extracted 4.2GB Data from GHarchive using Databrick Filesystem and performed analysis using pyspark, Spark API, SparkSQL, Spark dataframe
  - Exported aggregated and analysed data to AWS S3 bucket for storage and further analysis
- Environment: Python, Jupyter Notebook, TensorFlow, Keras, PCA, Deep Learning, AWS

## **Senior Software Engineer**

Mindtree Ltd - Bengaluru, Karnataka

August 2014 to December 2016

The project assignment was about analyzing general insurance dataset and provide important insight to client which can help the business improve their marketing strategy for business growth. The assignment also included automating a banking web application which would enroll clients/customers into system and provide various banking product to them with Esign Facility.

### **Responsibilities:**

- Extracted, transformed Insurance and banking data using SQL server, created complex queries using functions, view and subqueries
- Created interactive Tableau operational dashboards and stories, depicting the KPIs using calculated fields, groups, action, quick and context filter and hierarchies which helped client in improving marketing strategy-based customer segmentation resulted in 10% business growth
- Developed workbooks (with over 12 worksheets) with various functionalities and deployed them on server.
- Used functionalities like groups, sets, Actions (hyperlinks), context filters wherever required.
- Involved in Creation of tables, Manipulated the data using DDL and DML functions
- Trained and mentored 10+ team members which helped in meeting on-time project deadlines.

- Created 250+ automation test scripts and updated hybrid automation framework using HP UFT 11.5 and VB scripting, reported 50+ defects logged into HPQC, assigned to developers and conducted a walkthrough to explain the issues, retested, closed the defects and improved application functionality by 25%

Environment: Tableau, Window server, SQL Server, HP UTF, HP ALM, VB Scripting

## Education

---

### **M.S. in Business Analytics**

University of Texas at Dallas - Richardson, TX

### **B.Tech in Biotechnology**

SRM University - Chennai, Tamil Nadu

## Skills

---

- Programming: SQL, Python, R, Linux, Shell Script, SAS, Stata  
Data Visualization: Tableau 2020.1, Matplotlib and Seaborn in Python, GGPlot2 in R, Excel  
Big Data/Cloud: Hadoop, Hive, Spark, Pig, Sqoop, Map Reduce, Yarn, Kafka  
Cloud: AWS S3, EMR, EC2, Google Cloud Platform  
RDBMS: MS SQL Server, MySQL, MongoDB  
Libraries: NumPy, SciPy, Pandas, Scikit-Learn, TensorFlow, Keras, Pytorch, Spacy, NLTK, Vader  
Software/Tools: PyCharm, Jupyter Notebook, Visual Studio Code, HPQC, HP-UFT, R Studio  
Machine Learning: SVM, Regression, Classification, KNN, PCA, Natural Language Processing  
Statistical Model: Linear Model, ANOVA, t-test, chi-square, Hypothesis Testing