NAVYA SHIVA

(312) 918-9285 ● nshiva5@gmail.com ● linkedin.com/in/Navyas ● github.com/NavyaS ● Portfolio/Navyas ● Chicago, IL 60607

SUMMARY

Highly motivated Data science graduate having 3+ years' experience using predictive modelling, data mining and deep learning. Proficient in R, Python, SQL, Tableau, and Power BI using which I have created numerous machine learning models, optimized queries and built interactive dashboards to tackle several business problems. Experienced in creating Machine Learning Applications with ML using R, Python. Aspiring to bring actionable outputs using Data science skills to real-world industry problems.

EDUCATION

University of Illinois at Chicago	*Masters of Science in Business Analytics*	**(GPA:** 3.77/4)	**[Aug 2019 – Dec 2020]**
NIT Warangal, India	*Bachelor of Technology, Chemical Engineering* **(GPA:** 7/10)		**[Apr 2011 – Mar 2015]**

TECHNICAL SKILLS

Computer Programming : R, RShiny, Python(Pandas, Numpy, Scikit-learn, Pyspark, PyTorch), SQL, MATLAB
Software Tools : Tableau, MySQL, Jupyter, Big Data (Hadoop, Spark, MongoDB), Alteryx, Statistical R Programming, Stata, MS Excel
Analytics/Visualization : Machine Learning, Advanced Text Mining (NLP), Deep Learning, Tableau, Power BI, HealthCare Analytics, Marketing Analytics
ML Models Deployment : Amazon Web Services (AWS), Microsoft Azure ML, Amazon SageMaker, R Shiny, Tkinter, TabPy server

PROFESSIONAL EXPERIENCE

ML Research Assistant, *University of Illinois at Chicago, USA*   **[Oct 2019 – present]**
    Shopper Marketing Lift attribution [Inverse Reinforcement Learning| Optimization| Gurobi | Python| ETL]
- Built empirical model using Lasso regression with waiting effect, satiation effect of marketing tactic. Used Inverse Reinforcement learning as a constraint to attribute Lift to each tactic for each week. Obtained the path that led to maximum lift (3%) through optimization.

Data Scientist, *Merilytics, India*   **[Jun 2018 – Jun 2019]**
- Budget Forecasting- Based on trends and seasonality, developed a forecasting model (ARIMA) to predict the Net sales for 2019. Predicted with an accuracy of 92%. *[Python| Excel |Sales foresasting]*
- Built a demand forecast model using Random forest by combining the product features and customer segmented clusters generated using K-Means with an accuracy of 82%. *[Python| Scikit-Learn]*
- Automated the bi-weekly Net sales analysis report by data modelling to gather, transform, and aggregate the real-time global markets sales data using PySpark, that has reduced the cost of operation by $20K. *[PySpark |Tableau| SQL]*
- Product Profitability Analysis: Evaluated SKU level profitability and identified best products by region, and by category. Built interactive dashboard on Power BI which enabled better decision making of store operations and improved profits by 67%.

Machine Learning Intern, Mialo Technologies, *India*   **[Feb 2018 – May 2018]**
- Implemented a Computer vision algorithm to 'Automate License Plate Recognition', using OpenCV and Numpy libraries on Python in the Linux platform, predicted with an accuracy of 71% that enabled the business to automate the parking system.

Graduate Engineer Trainee, *Praj Industrues Limited, India*   **[Jul 2015 – Jul 2017]**
- Predicted the design parameters using Multivariate Regression Analysis with an accuracy of 93%, resulted in cost reduction of $106K *[R| Matlab]*

REAL TIME APPLICATIONS DEVELOPED

Twitter Topic Modeling Application [*Python| CorEx| Microsoft Azure ML Deployment*]
- Developed an application that applies topic modeling on the tweets of a subject using unsupervised LDA, and semi-supervised CorEx. Obtained the latent topics, analyzed their distribution and their sentiment over the weeks. Deployed the same in production using Azure ML deployment.

Chronic Kidney Disease Prediction [*R Programming | R Shiny | ggplot2*]
- Built an interactive screening tool that predicts risk of having a Chronic Kidney disease using Logistic regression model having 97% recall.

ACADEMIC PROJECTS

- **COVID-19 public response analysis:**
  Identified the major topics people discussed on twitter using Topic Modelling. Analyzed how the topics changed over time.
- **Health Information National Trends Survey** |Tool: **R**
  Predicted the significant variables impacting the usage of online medical records from the survey of questionnaire by conceptual analytical modelling. Reduced the variables from 500 to 30. Built logistic model to interpret the variable importance.
- **Prostate Cancer Analysis:**
  Based on the time to event, implemented **Survival Analysis** in **R** to predict the patient survival after 7 years from the diagnosis date. Built Cox proportional hazard models, to find hazard ratio for each therapy, stage, and age group.
- **Twitter Sentimental Analysis**:
  Built a logistic regression model from scratch and gradient descent to find the best parameters that fit the model. Performed binary sentiment classification to classify the tweet as good or bad by using TF-IDF as the feature. **[NLP, Text Mining]**
- **Recommendation system on MovieLens data**:
  Built recommendation model using Alternating Least Squares; **[Apache Spark | Python]**
- **Business Data Visualization, Delta Airlines** |Tool **[Tableau | R]**
  Visually represented and analyzed the factors that contribute to the performance of Delta Airlines. Captured the key competitor and major airports that needs to be focused to improve performance of Delta airlines