**SHIVAM THASSU**

Arlington, VA | [shivamt99@gmail.com |](mailto:shivamt99@gwu.edu) (703) 862-0058 | [www.linkedin.com/in/sthassu](http://www.linkedin.com/in/sthassu)

**PROFESSIONAL EXPERIENCE**



**Inovalon Inc. –** Bowie, MD

***▸ Data Science Innovation Lead*** *July 2019 – Present*

* **Spearheaded the ML/Analytics team of 5** to define problem statements, establish correlation, collaborate with engineering, business teams, stakeholders & customers, and developed key metrics, KPIs, optimizations to achieve business outcomes.
* Managed the **Innovation & research team** performing R&D into health data, developing ML/NLP models to tackle healthcare challenges, setting team priorities & business strategies, validating and executing algorithms and predictive models to investigate problems, generate features, detect patterns, recommend solutions, and **convey data stories to clients**.
* Performed ad hoc analyses of models & business performance trends and insights that **informed executive decisions**.

***▸ Data Scientist***

* **Aggregated unstructured data** from various sources to build an **NLP pipeline** encompassing several models for HEDIS supplemental **data extraction** (NER) of patient vitals, detailed medication & demographics from unstructured medical charts, which generated *~$350k* revenue.
* Built a **Hospital readmission** **predictive** model for patients who are diabetic, have heart & chronic diseases with an AUC of 84% using XGBoost (tested in simulated env on 2019 patient data with a potential cost savings of *~ $500k*+)
* Analyzed Health data - claims data, ICD codes, patient SDOH data to **uncover potential patients** to reach out who would most likely be interested in enrolling in a healthy meal plan (boosting client’s marketing & sales)
* Leveraged a **Sentiment model** to draw insights from audio data for the Call center team assisting them as a feature in improving their IHA completion rate by *25%.* (Google Speech-to-Text & VADER)
* Developed a **Heart** **stroke prediction model** using XBGclassifier for an external client to assist their Risk scoring strategy department.
* Built **Tableau dashboard** to visualize core business KPI saving *12hrs per week* of manual reporting.
* Created an **automated** performance reporting template, achieving an *80% reduction* in the manual preparation time.
* Performed ad hoc analyses of models & business performance trends and insights that **informed executive decisions**.

**GW Law Alumni and Development Office -** Washington, DC

***▸Data Science Analyst*** *June 2018 – Aug 2018*

* **Mined the data** from company databases to drive optimization & worked with data exploration, cleansing, manipulation, wrangling, and analysis of alumni donations data.
* Detailed **EDA and reporting** on alumni data to **uncover patterns** to help **predict alumni attrition** using LightGBM with *86.7 %* accuracy(meet theyearly donations target)

**Cognizant Technology Solutions -** Bangalore, India

***▸Programmer Analyst*** *(Domains: Healthcare, Telecom, Media)* *June 2013 – Feb 2017*

* Worked with **Data Engineers & ETL team** to support the processing, cleansing, and verification of the integrity of data.
* Interacted with **SQL databases** to create data models & Design/enhance dimensional data models.
* Part of **BI team, build reports, dashboards**, and tools to analyze reports, and present operations-related data using Tableau, Oracle OBIEE, Excel for project matrix. Improved the output by *30%* before the given deadline.

**EDUCATION**

| **The George Washington University -** Washington, DC | *Aug 2017 - May 2019* |
| --- | --- |
| *Master of Science - Data Science,**(GPA: 3.82/4.00)* |  |

* Area of focus: Data Mining & Analytics, Machine Learning, Data Visualization, Cloud Computing, Deep Learning, NLP

| **Shri Vaishnav Institute of Technology & Science** - Indore, India | *June 2008 - June 2012* |
| --- | --- |
| *Bachelor of Engineering - Computer Science*, *(GPA: 3.7/4.00)* |  |

**TECH STACK**



**Languages/ Scripting:** Python, SQL, Git

**Data Mining/Analysis:** Python (NumPy, Pandas, SciPy, Scikit-learn), PySpark

**Data Visualization:** Tableau, Power BI, Python (matplotlib, Plotly, seaborn, geoplotlib, ggplot2)

**Big Data/ Databases:** MySQL, Postgres, Oracle DB, Redshift, Snowflake

**Machine Learning:** Supervised & Unsupervised: Linear Regression, Logistic Regression, Decision Trees, SVM, K-NN Clustering, Bagging & Boosting, Ensemble models, Neural Networks, NLP

**Cloud Platforms:** Amazon Web Services (AWS EC2/S3), Google Cloud (GCP)