**Shivam Thassu**

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

**KEY SKILLS**

- Machine Learning

- Data Analytics

- Predictive Modeling

- Data Mining

- Natural Language Processing

- Data Visualization

- Statistical Analysis

- A|B Testing

- Data Strategy Implementation

- Project Management

- Research & Development

- Product Innovation

- Business Development

- Customer DS Consultancy

- Process Improvement

**PROFESSIONAL EXPERIENCE**

**Inovalon Inc**, Bowie, MD *July 2019 – Present*

Senior Data Scientist

*Role:**Technical Manager & Customer Consultant (TAM)*

* Managed, lead & executed initiatives in **identifying** **growth** **opportunities** for RWD Business Unit.
* **Oversaw Technical requirements**, trained customers to use the Data & addressed product-related technical queries.
* Contributed, impacted & **improved RWD Customer health** average score from *~52% to ~83%* (current) & ensured customer success & satisfaction.
* Designed, developed, documented **standards for data quality** and performed DQ checks for RWD customers; maintained data quality assurance throughout the Extract Enterprise Data Factory.
* **Designed Customer KPIs**, identified patterns & trends usecase for the internal & external teams.
* Collaborated with Engineering, Product, delivery, Ops & sales teams to improve process, identify gaps in the business & triage bugs, uncover inconsistencies in Product, ensure smooth delivery.
* Created & published well-structured playbooks & reporting with high quality research work **packaged as KB articles** & provided dedicated SME services for Advanced Analytics solution & consultation to key customers who contribute to *50% of RWD business (~$15 Mil)*

*Role: Data Science Innovation Lead*

* **Spearheaded the ML team of 3** 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 & internal executives**
* Performed ad hoc analyses of models & business performance trends and insights that **impacted executive decisions**.
* Collaborated & **lead partnership** with UMD Data Science PHD research team to tackle several research projects.

*Role: Data Scientist*

* **Aggregated unstructured data** from various sources to build a prototype **NLP pipeline** encompassing models for HEDIS supplemental **data extraction** (NER) of patient information from unstructured medical charts, which generated significant 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.

**GW Law Alumni & Development Tech**, Washington, DC*June 2018 – Aug 2018*

Data Science Intern

* **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*that helped them meet theyearly donations target.

**Cognizant Technology Solutions**, Bangalore, India*June 2013 – Feb 2017*

Programmer Analyst & Consultant

**Clients:** Telstra Corp Ltd (Telecom) | J&J (Healthcare) | Rodale Inc. (Media)

* Worked with **Data Engineers & ETL team** to support the processing, cleansing, and verification of the integrity of data. Successfully interpreted data to identify key metrics and draw conclusions. Proposed solutions to improve DQ efficiencies, leading to a 15% reduction in operating costs.
* Interacted with **databases** to create data models & Design/enhance dimensional data models.
* Analyzed data from *500k +* monthly active users and used outputs to guide marketing & product strategies, increased average engagement time *5x*.
* Part of the BI team**, building automated reports & dashboards**, and tools to analyze reports, and present operations-related data using Tableau, Oracle OBIEE, Excel for project matrix.

**ACTIVITIES**

* Winner of Inovalon 2021 Annual Hackathon. Bowie, MD - *Aug 2021*
* Developed *AI- Data Field Mapping Model “Try before you buy*” which reduced manual intervention for Data integration team to map and transform client data to Inovalon standard uniform data format before loading to IPORT-HD
* Model identified standard data field names for each 213 column which increased efficiency of new client onboarding, saving operations time from 4 hours to few minutes. [Video/Demo Link](https://inovalonglobal-my.sharepoint.com/:v:/r/personal/nag_manthena_inovalon_global/Documents/Recordings/Try%20before%20you%20buy%20Hackathon%20Demos.mp4?csf=1&web=1&e=E1Fa6P) [9:35 – 13:45]
* Winner of DC Data community Hackathon. Washington DC - *Jan 2019*
* Columbian College Outstanding Fellowship Award. Washington DC - *July 2017*
* Committee member of IEEE Computer Society Student Branch. India - *June 2012*

**TECH STACK**

|  |  |
| --- | --- |
| * **Scripting Languages** | Python, SQL, Git |
| * **Data Mining & Analysis** | Python (NumPy, Pandas, SciPy), PySpark |
| * **Databases** | MySQL, PostgreSQL, Oracle DB, Redshift, Snowflake |
| * **Frontend** | StreamLit |
| * **Machine Learning** | Supervised & Unsupervised ML: Linear Regression, Logistic Regression, Decision Trees, Support Vector Machines, KNN Clustering, Ensemble: Bagging & Boosting |
| * **AI & Deep Learning** | Neural Networks, CNN, Anomaly Detection |
| * **NLP** | SpaCy, NER, POS Tagging, Word embeddings |
| * **Data Visualization** | Tableau, Power BI, Python (Plotly, Seaborn, Bokeh) |
| * **Cloud Services** | Amazon Web Services (AWS), Google Cloud Platform (GCP) |

**EDUCATION**

**University of Maryland & NIH,** College Park, MD*July 2022 – Sept 2022*

Professional Certificate Program – AI Leadership in Healthcare

**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, 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)