Shantanu Solanki

101 E. Green Street, Apt 46, Champaign, IL

Mob. 217-200-2118 | Email | GitHub | LinkedIn | Portfolio Website

EDUCATION

University of Illinois, Urbana-Champaign, Illinois **Aug 2021 - May 2023**

Master of Science, Statistics **GPA 4.0/4.0**

Indian Institute of Technology, Dhanbad, India **July 2011 - May 2015**

Bachelor of Technology, Engineering **GPA 3.5/4.0**

RELEVANT COURSEWORK

Data Structures and Algorithms, Mathematical Statistics, Statistical Learning, Applied Regression and Design, Python for Data Science, Data Science Foundations, Database Systems, Natural Language Processing, Unsupervised Learning, Applied Bayesian Analysis

WORK EXPERIENCE

University of Illinois, Urbana-Champaign, Illinois **Aug 2022 - Present**

Research Assistant, ETC Lab, Department of Applied Health Sciences, College of Liberal Arts and Sciences

- Collaborated with a team of clinicians, health experts, programmers, etc. to create a master markdown file to transform raw survey data into actionable dataset, reducing time of preparation by **50%**.
- Presented SEM model findings to predict vital signs at several time points with/without specific interventions with **95%** accuracy to a diverse team of researchers.
- Co-ordinated with programmers to develop a React web application that registers survey responses of participants significantly cutting down survey time from **months** to **days**.

Ministry of Labor and Employment, Government of India, New Delhi, India **May 2018 - July 2021**

Data Scientist, Central Secretariat Services

- Unemployment Profiling on Labour Bureau's data (Employment Unemployment Survey, 2019):
  - Analyzed demographics data to predict unemployment trend at city, state and individual level using **Decision Trees**, **Logistic Regression**, **SVM**, etc. attaining an **AUROC** of **67%**. Presented the findings to the top management.
  - Identified the prominent factors for youth unemployment and resultant monetary deprivation using **logistic** and **linear** regression respectively, boosting the cause for targeted government spending which resulted in raising the overall employability of youth from **45.9%** in 2019 to **50.3%** in 2022.
- Worked on efficiently communicating data to the transfer committee of EPFO, a subordinate organization, using workforce dashboard created using **Power BI,** to make employee transfer procedures fair and transparent.

IITian's Hub (Higher Secondary Educational Institute), Dhanbad, India **June 2015 - March 2018**

Data Scientist

- Created a practice question recommendation system for students on the e-learning platform using **cosine similarity**.
- Developed a data-driven subscription prediction model for the marketing team, enabling targeted emails and offers that resulted in a significant **30%** revenue growth while optimizing costs.

ACADEMIC PROJECTS

Customer Churn Prediction: **R, Excel**

- Led a team of graduate students to predict customer churn with a **94% F1-score** using various supervised machine learning algorithms, effectively handling dataset imbalance with methods **like SMOTE, oversampling**, and **under-sampling**.

Customer Segmentation for Marketing Campaign: **Python**

- Collaborated with a team of three members to employ **K-Prototypes** and **Fuzzy c-means** for customer segmentation, achieving improved marketing efficiency and higher campaign success, highlighting data-driven optimization in marketing strategies.

e-Commerce Database Interface: **React, JS, MySQL**

- Developed an efficient e-commerce database interface using **SQL**, improved data management and reduced query response time, underscoring the impact of optimized data management in enhancing business operations.

Fashion Products Image Classification: **Python**

- Teamed up with colleagues to utilize various ML models to classify Fashion MNIST dataset, achieving **91.53% accuracy**, using ensemble models, showcasing potential to reduce manual labor in e-commerce categorization.

SKILLS

- **Data Analysis & Modeling:** Random Forest, XGBoost, Bootstrapping, Hypothesis Testing, PCA, Clustering, Generative Models, NLP, Time Series Modeling, Visualization, A/B Testing, Deep Learning, Recommendation System, Linear regression, GLMs
- **Programming**: Python (Pandas, Tensorflow, Keras, Sklearn, Matplotlib, Plotly, etc.), R, C++, SAS, JavaScript, Spark, Hadoop
- **Data Management & Visualization:** MySQL, MongoDB, Neo4j, Tableau, Power BI, GCP, Django, HTML, Git, AWS

HONORS AND ASSOCIATIONS

- Awarded the Dr. Art Kramer award by the Exercise Techhnology lab at the **UIUC** for outstanding contribution to the lab's research.
- Volunteered for the Non-Governmental Organization (Kartavya), working for the education of under-privileged kids in slums.