ANISHA SINHA

Bloomington, IN 📞 (930)215-7068 sinhaani@iu.edu 🔗 linkedin.com/in/anishasinha20

SKILLS

Languages: Python, R, SQL
Concepts: Machine Learning, A/B Testing, Statistical Inference, NLP
Tools: Powercenter Informatica, Tableau, Numpy, Pandas, Matplotlib, Plotly, MS Access, Oracle DB, Git

WORK EXPERIENCE

OVPUE IT, Indiana University - Bloomington   Oct '21 – current
Data Analyst | Python, SQL, Informatica

- Implemented ETL to university data from heterogeneous source systems into target database using Informatica.
- Created mappings using Designer and extracted data from various sources, transformed data according to the requirement.
- Developed the Informatica Mappings by usage of Aggregator, source filter usage in Source qualifiers and dataflow management.
- Built several reusable transformations and mapplets that were used in other mappings.

EDUCATION

Indiana University, Bloomington   Expected: May 2023
MS in Data Science   *Bloomington, Indiana*

K. J. Somaiya College of Engineering   Jun '17 – Jun '21
BE in Electronics and Telecommunications Engineering   *Mumbai, India*

PROJECTS

Analyzing radar echograms using Viterbi | *Python*   Dec '21

- Analyzed the effect of global warming on the north-south pole ice levels using radar echograms.
- Implemented simple bayes' net, Viterbi and HMM to examine the ice-rock and ice-air boundaries in the echograms.
- Compared and optimized the results of all the models to achieve an accuracy of 96%.

Financial Data Analysis using R and Machine Learning algorithms | *RStudio*   Mar '21

- Scraped financial data, performed data cleaning and EDA on the data to obtain better insight.
- Employed multiple linear regression and backward elimination method on the variables to make a parsimonious model for 5 Tata Group subsidiaries.
- Performed time-series analysis, employed ARIMA model on the 5 datasets, analysed its residual graphs using ACP and PACP plots.
- The goal was to compare the different models with respect to the economic rate of return for the year 2021 for each subsidiary.
- Achieved a 95 % accuracy for all companies.

POS Tagging using HMM and Gibb's Sampling | *Python*   Dec '21

- Built simple and complex Hidden Markov Models for POS tagging of 12 parts of speech.
- Implemented Gibb's Sampling for a bayes' net to to sample from the posterior distribution.
- Analyzed and compared three bayes' net models achieving an accuracy of 96%.

Bayesian Generalised Linear Models | *Python*   Sep '21

- Created a common model to handle Logistic, Ordinal and Poisson regression.
- Investigated Evidence Maximization and Cross validation for the model selection.
- Calculated the Mean Square Error to check the performance of the different models.

Sentimental Analysis using Naïve Bayes theorem | *Python*   Sep '21

- Implemented the Naïve Bayes algorithm with maximum likelihood and MAP solutions.
- Evaluated the algorithm using 10-fold cross validation.
- Computed improved accuracy's by varying the smoothing parameter.

CERTIFICATIONS

- **University of Michigan (Coursera)** - Python for Data Science
- **UC Davis (Coursera)** - SQL for Data Science
- **Duke University (Coursera)** - Statistics with R Specialization