Divya Sharma

M.Tech (Industrial & Management Engineering)

DATA SCIENCE INTERNSHIP AT HARVESTING

Email- sharmad@iitk.ac.in | Contact No- +918005851383

LinkedIn Id: https://www.linkedin.com/in/divyasharma3196/

| ACADEMIC DETAILS | | | | |
|------------------|--|---|------------|--|
| YEAR | QUALIFICATION | EDUCATIONAL INSTITUTION | PERCENTAGE | |
| 2019-21 | M.Tech (Industrial & Management Engineering) | Indian Institute Of Technology, Kanpur | 8.0 (CPI) | |
| 2013-17 | B.Tech (Mechanical Engineering) | University College Of Engineering, R.T.U., Kota | 69.65% | |
| 2013 | Class XII (CBSE) | Happy Public School, Alwar | 79.4% | |
| 2011 | Class X (CBSE) | St. Anselm's School, Alwar | 9.2 (CGPA) | |

BUILDING VALUATION

To determine monetary value of buildings in a specific area based on historical building data and remote sensing data

- Extracted amenity data in the region using QGIS(OSM) and used Haversine Formula to calculate distances between buildings and amenities
- Modelled the problem using Linear, Polynomial Regression, K-NN, Support Vector Machine, Decision Tree, Random Forest Regression
- Obtained R-squared value of 0.895 using Random Forest

FARM SCORE EVALUATION

To estimate the farm score of farms located in Bathinda region of Punjab using farm boundaries and other remote sensing data

- Estimated Farm Score using remote sensing data of farms in Bathinda, Punjab region
- Evaluated farm area, mean farm elevation, mean NDVI and farm distance from important amenities to calculate score
- Developed Farm Score formula using Analytical Hierarchy Processes

| ACADEMIC PROJE | CTS |
|--|--|
| Data Mining | Fake News Classifier - To build a system that identifies unreliable news articles using headings of news articles Vectorized headings of news articles using Count Vectorizer and TF IDF vectorizer Cleaned text using PorterStemmer, Stopwords class to improve the model accuracy Trained the dataset using Multinomial Naive Bayes(with and without Hyperparameter) and Passive Aggressive Algorithm Achieved an accuracy score of around 0.92 using Passive Aggressive Algorithm with TF IDF Vectorizer |
| | Medical Insurance Premium Prediction - To predict the cost of medical insurance premium and to study the factors affecting it Encoded the categorical data using OneHotEncoder and LabelEncoder and avoided Dummy Variable Trap Applied Simple and Multiple Linear Regression technique Incorporated an interaction feature BMI OF SMOKERS and significantly improved the Adjusted R- Square value to 0.82 |
| Statistical Modelling for Business | Bank Marketing Campaign Analysis – To extract information from existing marketing campaign to develop next marketing campaign Implemented Logit and Probit models to classify the subscription class Checked for multi collinearity of features using VIF and dropped down variables to avoid multi collinearity Achieved an accuracy of about 76%, precision of 0.73 and AUC of ROC curve, 0.85 |
| Analytics | Three Month Sales Prediction - To predict the three-month sales of 50 different items in 10 different stores Data consisted of daily sales of 50 different items in 10 stores from 2013 to 2017 Time series for Store-1 and Item-1 showed that series was seasonal and non-stationary Confirmed nonstationarity of series was by ADF and KPSS tests and seasonality was seen from ACF and PACF plots ARIMA did not perform well due to high frequency of data and NNETAR model was overfitting the series Prophet Model was the best model with SMAPE of 1.99 |
| Marketing Research | Brand Comparison between PUMA, ADIDAS and NIKE footwear - Conducted the online survey & did Analysis in SPSS Research questions - What role does price play while purchasing footwear? Which product is better on basis of build features? Is variety a vital factor while buying footwear? How do discounts offer work in simulating footwear purchase? What features of the current advertisement campaign are effective? What factors contribute to switching to another brand? Performed various parametric and non-parametric tests in SPSS & gathered customer insights |

ONLINE LEARNING AND CERTIFICATION

Machine Learning A-Z: Hands-on Python in Data Science | Machine Learning Practical | Python for Data Science by IBM | Python Data Structures by University of California | SQL for Data Science by University of California | Deep Learning A-Z: Hands-on neural networks

COURSEWORK AND SKILLS

Data Mining and Knowledge Discovery | Operations Research | Probability & Statistics | Introduction to Computing (JAVA) | Market Research | Stochastic Processes | Advanced Decision Models | Statistical Modelling for Business Analytics | Business to Business Marketing* | Game Theory*

Python(NumPy, Pandas, seaborn, sklearn, keras, TensorFlow) | R (Dplyr, Ggplot, Plotly, plm) | Java | SQL | MS Office

*Ongoing Courses

(May-June '20)

| POSITIONS OF RESPONSIBILITY | | |
|---------------------------------------|--|----------------------|
| Teaching Assistant, IIT Kanpur | Managed and provided support to 50 students and administered exams | Sept '2020 – Present |
| M.Tech Lab In Charge, IME, IIT Kanpur | Managed and assisted students in issues related to software and biometric authentication | 2019 - Present |
| Coordinator, LANX | Executed a college level LAN gaming tournament at LICE Kota | March 2017 |

ACHIEVEMENTS & CERTIFICATIONS

- Secured 1171 rank in GATE-2019 examination, conducted by Indian Institute of Technology, Madras
- Participated in Table Tennis competition at Aaghaz, IIT Kanpur and won silver
- Participated in group dance competition and won at Anukriti, UCE, Kota
- Participated in trekking expedition organized by Adventure Club, IIT Kanpur to Annapurna Base Camp, Nepal