SHIVAM SHARMA

M.Tech (Industrial & Management Engineering)

✉ shivsha@iitk.ac.in   📞 +918954437753
in linkedin.com/in/shivam-sharma-258728195/

ACADEMIC DETAILS

YEAR	QUALIFICATION	EDUCATIONAL INSTITUTION	CPI/PERCENTAGE
2019-21	M.Tech(Industrial & Management Engineering)	Indian Institute Of Technology, Kanpur	7.76*(CPI)
2013-17	B.Tech (Mechanical Engineering)	Dayalbagh Educational Institute, Agra	7.29(CPI)
2012	Class XII (CBSE)	Kendriya Vidhyalaya No.3, Agra	85.20%
2010	Class X (CBSE)	Dayanand Bal Mandir School, Agra	9.0

upto 2nd semester

ACADEMIC PROJECTS

Data Mining and Knowledge Discovery

Quora Question pair similarity problem *(Sept'19-Nov'19)*

- The data consisted of 404290 question pairs. 790 Features were created by Exploratory Data Analysis of Question pairs.
- Steps include **Basic Statistics**, **Simple Feature extraction** by **Intuition**, **Text Preprocessing**, **Advanced feature extraction** using **Fuzzy features**, Analysis of text with **word cloud**, **Data Visualization** using **T-SNE**, **Featurizing** text data with **tf-idf weighted Word Vectors**.
- Applied Random model , **Logistic Regression** ,**Support vector Machine (SVM), XGBoost** with **Hyper parameter Tuning**
- **XGBoost** showed the best results with **train log loss .345**, **test log loss .357** with Precision = **.833** and Recall **.9**

Statistical Modeling For Business Analytics

Analysis of the Factors Affecting Sales Price of house in King Country, USA *(Jan'20-Mar'20)*

- The dataset contains **19** house features including the **price** (dependent variable),along with **21613** observations.
- Carried out **multivariate** statistical **regression analysis** to study the factors influencing house prices, performed EDA, calculated correlation matrix and checked for **Multi-collinearity** and **Omitted variable bias.**
- **Breusch-Pagan** test showed heteroskedasticity, hence "**heteroskedastic robust errors**" were used.
- Built **linear regression** models, Used **Backward elimination approach** in multiple regression model and finalized a model with **Adjusted R2** without and with robust error was 0.676 & 0.694 respectively.

Telecom Customer Churn Prediction *(Mar'20-May'20)*

- The dataset contains 7043 rows (customers) and 21 features Such as" tenure", "online security", "paperless billing" etc.
- Performed EDA, Correlation matrix, applied **SMOTE** to balance the data **and RFE (Recursive Feature Elimination)** to select the 15 significant features in the baseline model.
- **Logit** and **Probit** models were used for classifying the Churn class, features were dropped on the basis of **p-value** and **VIF.**
- Reported an **accuracy** of about **79%** ,**precision** of **73.8%** and a **recall** of **62.4%**,**AUC** of **ROC curve** was **0.83**

Stochastic Processes

Credit Card Fraud Detection Using Hidden Markov Models *(Mar'20-May'20)*

- Used **simulated** data of credit card user to train a **Hidden Markov Model** and estimated **transition probabilities** and **emission probabilities** using **Forward-Backward algorithm** and sequentially predicted whether the upcoming transaction is fraud or not with **Recall 81% and F1-score 0.67**.

Marketing Research

Measuring loyalty of the viewers towards the TV channels and identifying audience profile of TV channels with highest viewer loyalty *(Jan'20-May'20)*

- The purpose of this Marketing research study is to find out that with the introduction of new TRAI rules and new price segmentation which customer will remain loyal to which channel.
- Designed **cross sectional one shot case study** dynamic survey form using **Scaling techniques, pretesting** to control internal and external validity of Research Design. Data collected using online survey, Focus groups and Personal Interview. Conducted Exploratory, **Descriptive Research in SPSS** using primary data obtained by convenience random sampling. Analyzed sampled data using statistical test **(One Sample t-test, One way ANOVA, independent t-test)** to test our hypothesis.

SUMMER INTERNSHIP

Data Science Intern at Harvesting Farmer Network *(May'20-Jun'20)*

Objective : Twitter match making (tag the potential buyer to seller) and Tag the officials

- Applied an algorithm to scrape tweets of targeted username **@HarvestingFN** from start date to end date of our choice using **twitterscrapper.**
- Done tweet text pre-processing such as stop words removal, conversion to lowercase using **NLTK,** pandas.
- Extract the followers list that follow HFN(@HarvestingFN) twitter handle by using **tweepy cursors** .
- From the list of followers extract their location using **user.location() function , Geolocation and Geocoding API**.
- Applied an algorithm to tag the potential buyers and officials on the respective tweet & used **api.update_status(status)** function for this purpose.

COURSEWORK AND SKILLS

Relevant Courses	**Data Mining** and Knowledge Discovery | | Statistical Modeling for **Business Analytics**| **Probability & Statistics** |Introduction to **Stochastic Process** | **Marketing Research** |Accounting & Finance | Introduction to computing |Operation Research
Technical Skills	**Python**(NumPy, Pandas, Matplotlib, Scikit-learn) | **R** | **SQL** | **CATIA**(Computer Aided Three Dimensional Interactive Application) |

POSITIONS OF RESPONSIBILITY

Student Nominee SPGC, IIT Kanpur (Senate Post Graduate Committee)	Student Representative for matters concerning PG Academics.	*(Sep'19-Present)*
Placement coordinator D.E.I ,Agra	Responsible for conducting placement exams and interviews by coordinating with the firms.	*(Aug'16-Mar'17)*

ACADEMIC ACHIEVEMENTS & CERTIFICATIONS

- Awarded certificate for Successful completion of **National Service Scheme (NSS)** Programme (240 hrs) in the tenure 2013-15.
- Awarded Certificate for actively participating in **community service and social service** (240 hrs NSS camp in 2013).
- Awarded Certificate for successful completion of course on **CATIA**(Computer Aided Three Dimensional Interactive Application) *('Jul15-Sep'15)*
- Awarded Certificate for successful completion of course **SQL for Data Science**. *('Jul-2020)*