ABHINAV PATERIA

pateria@iitk.ac.in | +919005827019
MTech (Industrial & Management Engineering) https://www.linkedin.com/in/abhinav-pateria-14a170181

ACADEMIC DETAILS

YEAR	QUALIFICATION	EDUCATIONAL INSTITUTION	PERCENTAGE
2019-21	M.Tech(Industrial & Management Engineering)	Indian Institute Of Technology, Kanpur	8.49*(CPI)
2013-17	B.Tech (Mechanical Engineering)	Gurukula Kangri Vishwavidyala, Haridwar	83.86%
2012	Class XII	Sarawati Vidya Mandir Inter College Hamirpur UP	84.40%
2010	Class X	Sarawati Vidya Mandir Inter College Hamirpur UP	78.66%

**upto2nd semester*

ACADEMIC PROJECTS

Applied Machine Learning

Quora Question pair similarity problem *(Feb'20-Apr'20)*

- 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 of **.833** and Recall of **.9**

Statistical Modeling for Business Analytics

Analysis of the Factors Affecting Sales Price of house in King County, 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**, performed EDA, checked **Multicollinearity** by VIF & correlation matrix
- **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**
- Logit showed better results as **accuracy** of about **79%** ,**precision** of **73.8%** and a **recall** of **62.4%**,**AUC** of **ROC curve** was **0.83**

Data Mining and Knowledge Discovery

Box office Movie's worldwide Revenue Prediction *(Sept'19-Nov'19)*

- The data consisted 3000 train data points, 4398 test data points with 22 attributes and revenue as target variable
- Steps include **Exploratory data Analysis,** Feature **Engineering**, treating missing values
- Applied **Linear Regression, Random Forest**, **XGBoost** . **XGBoost** showed the best results with **RMSE value 2.02**

COURSEWORK AND SKILLS

Relevant Courses	Data Mining and Knowledge Discovery | Applied Machine Learning| Statistical Modeling for Business Analytics| Advance Statistical Methods for Business Analytics |Probability & Statistics |Introduction to Stochastic Process | Operations Management Operations Research for Management |Introduction to Computing
Technical Skills	Python(NumPy ,Pandas, Matplotlib, Scikit-learn)| R|SQL| MS Office

SUMMER INTERNSHIP

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

Objective : Twitter matching 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 as stop words removal, conversion to lowercase using NLTK , pandas.
- Developed algorithm to extract the information of city, state of the seller from tweet using Python codes.
- Extract the followers list that follow HFN twitter handle by using tweepy cursors
- From the list of followers extract their location using **Geolocation and Geocoding API**.
- Developed an algorithm to match the city of seller with followers and officials.
- Applied an algorithm to tag the potential buyer and officials on original tweet using update_status function.

POSITIONS OF RESPONSIBILITY

- **PG Secretary, Academics & Career Council, IIT Kanpur** *Responsible for managing various sessions of council and gathering students opinion on academics.*
- **Internship Coordinator at IME department, IIT Kanpur** *Coordinated in a team of 4 and completed the Summer Internship process for IME-MTech batch.*
- **Student Senate Nominee to DPGC , IIT Kanpur** *Student representative on issues related to coursework,thesis work and coordination with faculties.*
- **Orientation Team Member, Counseling Service,IITK** *Coordinated with core team of counseling service and helped the students during winter admission.*
- **Branch Representative**, Mechanical Engineering, GKV Haridwar.

ACADEMIC ACHIEVEMENTS & CERTIFICATIONS

- **R-Programming** A-Z: R for Data Science with real exercises at **Udemy**
- **Python for Data Science** and Machine Learning Bootcamp at **Udemy**
- Secured **99.59 Percentile** in **GATE Mechanical Examination.**
- Secured **1st position** in **B.Tech.**
- Won 2nd price in 'Concatenation' and 'Junkyard' event in JNAGNI 2015, GKV Haridwar.