VIPUL KUMAR
M.Tech.(Industrial & Management Engineering)

🔗 https://www.linkedin.com/in/vipul54kumar/
✉ vipulk@iitk.ac.in 📞 +91-7683093810

Year	Qualification	Educational Institute	Percentage
2019-21	M.Tech.(Industrial & Management Engineering)	Indian Institute of Technology, Kanpur	7.49* (CPI)
2013-17	B.Tech. (Mechanical Engineering)	BIET Jhansi	75.86%
2013	Class XII, (CBSE)	P.M.S. Dhampur, UP	86%
2011	Class X,(CBSE)	P.M.S. Dhampur, UP	9.8(CGPA)

upto 2nd semester

INTERNSHIP

Data Science Intern at Harvesting, Chandigarh   **Mentor:** Mr. Ishaan Kochhar   (*April-June 2020*)

- **Title:** Transportation Management of Farmers Products.
- This project aimed to determine the transportation cost required for a user to buy goods from a selected farmer.
- Used **Geolocation and Geocoding API** for user location identification and then used Place API to detect nearest railway stations for that location.
- Performed Web Scraping with **Selenium** to get the shortest distance between the two locations from the Indian railway's official website.
- Handled **Indian railway parcel rate dataset** with Pandas to determine the transportation cost for the input conditions. This process was made completely automatic in Python and was ready to run in the background of the company's official website.
- Used **Zapier** for bringing automation in **Airtable** to facilitate data storage.
- Performed Data Cleaning with **pandas** to apply geocoding on user-inputted addresses.
- Used python library **Requests** to access data and worked with HTTP methods GET and POST.
- Tools used: Pandas, Numpy, Selenium ,Google Maps API, requests, Airtable, Zapier.

ACADEMIC PROJECTS

Data Mining	**Predicting box office revenue of a movie using Random Forest** (*Oct'19-Nov'19*) • This Project predicted how much revenue a movie is going to make at the box office. • Steps include **data pre-processing, exploratory data analysis, Linear Regression, Random Forest** models building and finally predicting test data from finalised random forest model. • For movie revenue top five important variables came out be popularity, budget-year-ratio, weekday-release, month of release, and main-genre. • Packages: plotly, ggthemes, dplyr, stringr, ggplot2, knitr, viridis, VIM, lubridate, RandomForest.
Statistical Modelling for Business Analytics	**Prediction Modelling of bank loans using Logistic Regression using a bank Dataset** (*Mar'20-Apr,20*) • Objective: To classify whether a loan applicant will be a defaulter at a later stage or not based on factors such as credit amount, employment, property, purpose, age, housing, dependents, credit cards, etc. • Class credit rating was **unbalanced** with 70% of dataset belonging to 'good' class. • It was observed that offers for Car loan can pick up more customers for loan from the bank. **Logistic Regression Models** were used to classify the credit rating class. Final Model gave an **accuracy of about 76% and a precision of 74% and recall of 63%, AUC of ROC curve was 0.87** which shows that model predictive power is good. • Packages used: numpy, pandas, matplotlib, seaborn, sklearn.
Applied Machine Learning	**Movie Review Sentiment Analysis** (*Feb'20-Apr'20*) • Objective: To predict the sentiment (Negative, Somewhat Negative, Neutral, Somewhat Positive, Positive) of Rotten Tomatoes movie based having 1.5 lakh reviews. • Performed data-cleaning and pre-processing including **Exploratory Data Analysis, Feature Engineering, Data Visualization** including **word cloud** for each sentiment. • Feature extraction techniques like **CountVectorizer, TF-IDF(Term Frequency- Inverse Document Frequency).** • Generated **classification report** and **confusion matrix** using **Logistic Regression**, **Stochastic Gradient Descent, Random Forest**.Random Forest with TF-IDF was observed as best model with **accuracy of 0.63.**

COURSEWORK AND SKILLS

Relevant Courses	**Applied Machine Learning	Probability and Statistics**	**Data Mining and Knowledge Discovery**	Operations Research for Management	Introduction to Computing(JAVA)	Marketing Research	**Statistical Modelling for Business Analytics** |Finance and Accounting(ongoing)
Technical Skills	**Python**(numpy,pandas, matplotlib, seaborn, scikit-learn, plotly, selenium) || **R**(ggplot2, dplyr, stringr, lubridate, plotly, ggthemes, knitr, rmarkdown) || **SQL** || JAVA || MS Office(Excel, Word, PowerPoint)

ONLINE LEARNING & CERTIFICATIONS

- Certified Python Financial Analytics course from Henry Harvin.
- SQL-**MySQL** for Data Analytics and Business Intelligence.
- Natural Language Processing in Python.

POSITION OF RESPONSIBILITIES

- Orientation Team Member & Buddy at **IITK Counselling Service**.
- Core Team Member for Open House, an event organized at IIT Kanpur Diamond Jubilee celebration.

ACHIEVEMENTS

- Secured All India Rank 785 in GATE 2018.