ABHISHEK RAI
MTech (Industrial & Management Engineering)

✉ **abhinitp@iitk.ac.in** | ☎ +919660246643
🔗 https://www.linkedin.com/in?/abhishek-rai-aba43a1a2

ACADEMIC DETAILS

YEAR	QUALIFICATION	EDUCATIONAL INSTITUTION	PERCENTAGE
2019-21	M.Tech(Industrial & Management Engineering)	Indian Institute Of Technology, Kanpur	7.78* (CPI)
2014-18	B.Tech (Mechanical Engineering)	National Institute Of Technology, Patna	8.14
2013	Class XII (RBSE)	Mayank Senior Sec. School ,Kota	81.8%
2011	Class X (RBSE)	Mayank Senior Sec. School ,Kota	78.83%

upto2nd semester

SUMMER INTERNSHIP

Data Science Intern at Harvesting Farmer Network   (May'20-Jun'20)
- Objective of the project was to create a **Chatbot** to communicate with farmers and deploy it on **Whatsapp**
- Used feature of **Dialogflow** to created chatbot like Agent creation, intent, entity, training, integration, fulfillment etc
- Trained bot by giving every possible user expression and Fetch the details of farmers
- Stored the data into the **Airtable** by using **Webhook(Json)** to Dialogflow
- Connected the **Airtable** with twitter account (company) using **Zapier** and deployed Chatbot in whatsapp using **Twilio**

ACADEMIC PROJECTS

Data Mining	**New York City taxi Demand prediction** (Sep'19-Nov'19) - Predicted the number of pickups in a given region in New York city from the given dataset of 19 features - Steps included Data visualization, Data preparation using **k Means Clustering** and **Time Binning** - Models used were simple moving average**,** weighted moving average**, Exponential weighted moving averages** - Regression models used are Linear Regression and ensemble methods like **Random forest Regressor** and **Boosting** - Models were tuned and compared on **MAPE** metric and **MAPE** was .1293 from **Exponential weighted moving averages**
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 - Determined correlation matrix and checked for **Multicollinearity** and performed EDA - **Breusch-Pagan** test showed heteroskedasticity, hence "**Heteroskedastic robust errors**" were used - **Adjusted R2** with and without robust error was 0.694 & 0.676 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 - Skewed dataset with 21 variables, applied **SMOTE and RFE (Recursive Feature Elimination)** to improve the baseline model - **Logit** and **Probit** models were used for classifying the churn class - Reported an **accuracy** of **Logit** about **79%** ,**Precision** of **73.8%** and a **Recall** of **62.4%**,**AUC** of **ROC curve** was **0.83**
Applied Machine Learning	**Netflix Recommendation System** (Feb'20-May'20) - The dataset contains Movies id followed by User id, date and Ratings are on a scale from (1 to 5) - Performed preliminary data analysis and **time based splitting** and created **sparse** matrix from data frame - From sample training data created **13 initial features** for r**egression** then trained it with a **Xgboost regressor** - Applied different models such as **Surprise baseline model, Surprise Knn model, Matrix factorization techniques** - Applied **Xgboost** with surprise models, compared on **RMSE** and **least RMSE** (1.0726) from **Matrix Factorization(SVD)** **Statistical Analysis on factor influencing Life Expectancy** (May'20-Jun'20) - The dataset consisted 193 countries from year 2000-2015, 22 Features such as **Economical factors** and **Social factors** - Checked **Heterogeneity** across countries (or entities) and years - Panel Models used were pooled regression**, Entity& Time Fixed effects regression and** Random Effects regression - The highest Adjusted R-squared was 0.9482 with **Binary** Regressor model for entities - Conducted **Hausman** test and tests for checking **heteroskedasticity**, **Serial Collinearity** and **Panel effect**

COURSEWORK AND SKILLS

Relevant Courses	**Data Mining and Knowledge Discovery** | **Statistical Modeling for Business Analytics**| **Probability & Statistics** | **Applied Machine Learning**| Operations Management | Business Management using Cloud |E- Supply Chain Management | Operations Research for Management
Technical Skills	Python(NumPy, Pandas, Seaborn, Scikit-learn)| R(Dplyr, Ggplot2, Plotly, plm)|SQL| MS Office

POSITIONS OF RESPONSIBILITY

- **Alumni Relations Coordinator IME Department:** To maintain alumni database & to arrange webinars
- Responsible for the activities in TECHNO CULTURAL FESTIVAL **CORONA** organized by our students's union in **NIT PATNA**

ACHIEVEMENTS & CERTIFICATIONS

- Got appreciationcertificate from**TATA MOTORS** for AUTOMOBILE BASICS AND ADVANCED SYSTEM with A++
- R-Programming A-Z: **R for Data Science** with real exercises at **Udemy**
- **Introduction to Python** Programming for Data Science at **Udemy**
- Secured **1216 Rank** in GATE Mechanical 2019,conducted by Indian Institute of Technology, Madras
- **Acted** as a Hike Sweeper in a team of 23 members from the **Adventure Club** of IIT Kanpur to **Annapurna Base Camp, Nepal**(Oct'19)
- **Teaching Assistant in IME Department**: Managed and provided support to 20 students and administered exams