

# VIVEK PRAJAPAT

Email: [pvivek@iitk.ac.in](mailto:pvivek@iitk.ac.in) | Contact No: +91-9685760268

MTech (Industrial and Management Engineering)

LinkedIn: <https://www.linkedin.com/in/vivek-prajapat>

ACADEMIC DETAILS			
YEAR	DEGREE	INSTITUTE	CPI/%
2019-Cont.	M.Tech. (Industrial & Management Engineering)	Indian Institute of Technology, Kanpur	9.24*
2013-17	B.Tech. (Civil Engineering)	National Institute of Technology, Bhopal	7.39
2013	Class – 12 <sup>th</sup>	Adarsh High Senior Secondary School, Bikaner	84.40%
2011	Class – 10 <sup>th</sup>	Apex Public Senior Secondary School, Jodhpur	83.67%

\* Up to 2<sup>nd</sup> Semester

INTERNSHIP	
<b>Data Science Intern, Harvesting</b>	(Apr'20 – Jun'20)
<b>Building Price Valuation</b>	
<ul style="list-style-type: none"><li>Estimated valuation of buildings based on characteristics of building and remote sensing data, <b>Amenities</b> data extracted with the help of <b>QGIS</b>.</li><li>Performed EDA used Geospatial Analysis, calculated Haversine distance between building and amenities.</li><li>Important <b>inferences</b> were made based on testing some basics intuition/<b>hypothesis</b>, statistically and by visualizations.</li><li>Modelled using <b>Linear, Polynomial regression, Regularization(L1&amp;L2), K-NN, Support Vector, Decision Tree, Random Forest regressor</b> and did <b>cross validation</b> to tune <b>Hyperparameters</b>. Best Models were <b>Polynomial Regression(d=2)</b> with <b>R<sup>2</sup> 0.895</b> and <b>Random Forest Regressor</b> with <b>R<sup>2</sup> 0.9</b>.</li></ul>	
<b>Farm Score</b> – Designed and calculated Farm Score for different Farms based on remote sensing data. For calculation <b>Farm area</b> , mean <b>elevation</b> of farm, mean <b>NDVI</b> and <b>distance</b> from important <b>amenities</b> were used. For designing formula of Farm Score, <b>AHP</b> (Analytic hierarchy process) was used.	

ACADEMIC PROJECTS	
<b>Amazon Fine Food Review Classification</b> (Data Mining and Knowledge Discovery)	(Aug'19-Sep'19)
<ul style="list-style-type: none"><li>Classified sentiment based on review text, performed data cleaning and pre-processing by <b>Stemming</b>, stop-word removal, and <b>Lemmatization</b>.</li><li><b>Class imbalanced</b> problem handled by <b>under sampling</b> and for train-test splitting <b>time-based sampling</b> was used.</li><li>Feature extracted using <b>Bag of Words, TF-IDF, Average Word2Vec, TF-IDF Word2Vec</b>. Applied <b>Logistic Regression, Random Forest Classification, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors</b> with Hyperparameter tuning.</li><li>Used <b>Accuracy, Precision, Recall and F1-Score</b> as metrics for comparison. Best Model was <b>Logistic Regression</b> with <b>TF-IDF</b> having <b>accuracy 0.93</b>.</li></ul>	
<b>Analysis of Factors affecting car price</b> (Statistical Modelling for Business Analytics)	(Jan'20-Feb'20)
<ul style="list-style-type: none"><li>Carried out <b>multivariate statistical regression</b> analysis to study which variables are significant in predicting the price of a car.</li><li>Performed <b>EDA</b>, calculated <b>measure of fit, correlation matrix</b>, performed <b>Breusch-Pegan</b> test for <b>heteroskedasticity</b>, checked for <b>multicollinearity</b> using <b>VIF</b> (Variance Inflation factor) and looked for <b>omitted variable bias</b>.</li><li>Feature elimination is done using <b>RFE</b> (Recursive Feature Elimination) based on <b>p-value</b> and finalized model with <b>R<sup>2</sup> 0.918, Adjusted R<sup>2</sup> 0.915</b>.</li></ul>	
<b>Store Item Demand Forecasting</b> (Statistical Modelling for Business Analytics)	(Mar'20-June'20)
<ul style="list-style-type: none"><li>Predicted 3 months of sales for 50 different items in 10 different stores from past 5-year sales data using <b>time series</b> techniques.</li><li>Checked for <b>stationarity, trend, seasonality</b> using <b>ADF-test</b> (Augmented Dickey-Fuller), <b>KPSS-test, Decompose Plot</b>.</li><li><b>AR, MA, ARMA, ARIMA</b> time series models applied. For <b>ARIMA(p,d,q)</b> parameters p and q, <b>PACF</b> (Partial Autocorrelation function) and <b>ACF</b> used.</li><li>Time series made <b>stationary</b> by <b>Differencing(d)</b>. <b>RMSE</b> and <b>MAPE</b> used for evaluation metric.</li></ul>	
<b>Customer Segmentation to define marketing strategy</b> (Applied Machine Learning)	(Jan'20-Feb'20)
<ul style="list-style-type: none"><li>Data visualization &amp; exploration by checking missing data, data duplication, plotted <b>KDE</b> (Kernel Density Estimate) and heatmap of <b>correlation metrics</b>.</li><li><b>K-mean clustering</b> applied. Optimal numbers of clusters are found by <b>Elbow method</b>. Visualization of clusters is done using <b>PCA</b> (Principal Component Analysis). Dimensionality reduction is done using <b>Autoencoders</b> and the same process is applied on reduced features.</li></ul>	
<b>New York Taxi Price Prediction</b> (Data Mining and Knowledge Discovery)	(Oct'19-Nov'19)
<ul style="list-style-type: none"><li>Predicted fare of a Taxi trip from 55M taxi trips data consist of Fare, trip duration, passenger count, etc. <b>Dask</b> was used for working on huge dataset.</li><li>Performed Data cleaning, exploration, pre-processing. Tested some basics intuition/<b>hypothesis</b> statistically and by visualizations.</li><li><b>Linear, Polynomial Regression, Random Forest, Light GBM</b> applied. After featurization and tuning, <b>Light GBM</b> performed best with <b>RMSE 3.63</b>.</li></ul>	

COURSEWORK AND SKILLS	
<b>Academic Courses</b>	<b>Data Mining   Probability &amp; Statistics   Applied Machine Learning   Statistical Modelling for Business Analytics   Introduction to Computing (JAVA)   Stochastic Process   Simulation of Business Systems   Advanced Statistical Methods for Business Analytics   Operation Research</b>
<b>Skills</b>	<b>Machine Learning   Statistical Analysis   Natural Language Processing   Deep Learning (Beginner)   PYTHON (Numpy, Pandas, Scikit-Learn, Matplotlib, Seaborn, Statmodels, Geopandas, Folium)   R   SQL   JAVA   MS-Office etc.</b>

ONLINE COURSES	
<b>IBM Data Science Professional Certificate   Deep Learning Specialization by deeplearning.ai   Applied Data Science Specialization   SQL for Data Science Market Research Specialization by UC Davis   Machine Learning by Stanford University   Python for Data Science and Machine Learning Bootcamp</b>	

POSITION OF RESPONSIBILITY	
<ul style="list-style-type: none"><li><b>Students' Senate Nominee to Academic Senate Standing Committee (Senate Education Policy Committee)</b></li></ul>	(Jun'20 - Present)
<ul style="list-style-type: none"><li><b>PG Senator Y19, Students' Senate</b> – responsible for bringing issues faced by PG students to the Student' Senate.</li></ul>	(Aug'19 - May'20)
<ul style="list-style-type: none"><li><b>Alumni Relations Coordinator</b>, IME Department, IIT Kanpur.</li></ul>	(Aug'19 - Present)
<ul style="list-style-type: none"><li><b>Orientation Team Member</b>, Counseling Service, IIT Kanpur.</li></ul>	(Dec'19 - Jan'20)
<ul style="list-style-type: none"><li><b>Senior Executive, Raktarpan (Blood Connect)</b>.</li></ul>	(Aug'19 - Present)

ACHIEVEMENT AND EXTRA CURRICULAR ACTIVITIES	
<ul style="list-style-type: none"><li><b>Academic Excellence Award</b>, 2019. Recognition for outstanding academic performance across all departments.</li><li>Secured <b>AIR 931</b> with a score of 730/1000 in <b>GATE2019</b>.</li><li>Awarded Central sector scheme of scholarship by <b>MHRD</b> for <b>top 1 percentile</b> students of HSC examination.</li><li>Awarded <b>NCC, 'B'</b> and <b>'C'</b> certificate with <b>A grade</b>.</li></ul>	