SHIVAM KUMAR TYAGI
M.Tech (Industrial & Management Engineering)

tshiv@iitk.ac.in |+919773840741

Year	Qualification	Educational Institution	Percentage
2018-20	M.Tech (Industrial and Management Engineering)	Indian Institute Of Technology, Kanpur	7.81* (CPI)
2013-17	B.Tech (Mechanical Engineering)	Harcourt Butler Technical University, Kanpur	74.08%
2012	Class XII (CBSE)	C.C.D.P.S. Ghaziabad, U.P.	87.2
2010	Class X (CBSE)	C.C.D.P.S. Ghaziabad, U.P.	85.5

upto 2nd semester

INTERNSHIP

Data Science Intern at Bayer Crop Science (Monsanto Holdings Pvt. Ltd.), Bengaluru *(May-July'19)*

- Title: Visualization and Analysis of Stewarded Insect Control Data to enable quick and informed decisions for Insect control team in St. Louis
- Data exploration & manipulation: Exploration of 5 years insect assay data including metadata areas such as plant, insect, protein names, assay type and results (in mortality % or plant damage rates upon infestation by insects), Data having mean values were extracted and handled discrepancies in data
- Built a Visualization tool (R Shiny Dashboard) with **real time data: Interactive web app (Dynamic UI)** using R Shiny was built for insect control assay data coming from in-planta studies and field assays. It involved understanding the data, writing codes for server & user interface components following **data science best practices** including version control of the codes (**GitHub**) and built my own packages for shiny app
- **Published Shiny dashboards** on an enterprise server and demonstration of usage: This step involved publishing the R Shiny dashboard on the *in house* Science-at-Scale server and demoing the tool usage in the Data Strategy meeting. Packages frequently used: ggplot2, plotly, data.table, rattle, RGtK2, shiny, shinyWidgets, shinydashBoard, dplyr, shinyjs, tidyr, lubridate, rmarkdown, knitr, RColorBrewer, roxygen2, plumber

ACADEMIC PROJECTS

Data Mining	**Music Recommendation System:** To predict the chances of a user listening to a song repetitively after first observable listening event • Built Music Recommendation System using a dataset that holds Asia-Pop music library with over **30 million tracks** • Stratified Sampling is done on 70 lakhs tuples, Performed pre-processing on the dataset including **PCA,** using **K-fold cross-validation** for training and validation dataset, Generated classification report & confusion matrix. Fitted the training data in model built using classifiers like SVM, decision tree , KNN and XGBoost, Got best accuracy of 69.62% with XGBoost • Packages used – ggplot2, dplyr, data.table, tibble, Caret, Xgboost, Gbm, Ranger
Statistical Modelling for Business Analytics	**Predicting quality of RED WINE using Statistical Regression Models** • The data consisted of 1599 observations of 12 variables, Determined correlation matrix and checked for **Multicollinearity** • Carried out multiple linear regression with quality as the dependent variable and others as independent, found "citric.acid", "residual.sugar" and "fixed.acidity" were statistically insignificant hence removed them from the model. **Breusch-Pagan test** showed heteroskedasticity in the data, hence **"heteroskedastic robust errors"** were used for regression on significant variables using **"lmrob()"** function from **"robustbase"** package. Adjusted R2 without robust error and with robust error was **0.3567** & **0.381** respectively. • Statistically significant variables were alcohol, volatile acidity, density ,chlorides and pH **Predicting Income class using Logistic Regression using Adult data set from UCI Machine Learning Repository** • Objective: to predict whether a person's income is <50K or >=50K (target variable) based on factors such as "age", "education", "marital status", "gender", "income" etc. Income class was unbalanced with about 75% of data belonging to class <50K. (46,033 observations) • It was also observed that as age increases the no of working hours for people belonging to <50K class first increases then decreases but for people belonging to >= 50K class it remains constant and then starts decreasing • **Logit and Probit** models were used for classifying the income class. The performance was similar to an **accuracy of about 84.3% and a precision of 61.9% and a recall of 52.8%, AUC of ROC curve was 0.88** which shows that the model's predictive power is good. Packages used: ggplot2, GGally, dplyr, ROCR, data.table, plotly
Marketing Research	**Consumer behavior towards Adidas footwear:** Conducted the online survey & did Analysis in SPSS • Research questions -Is Price or Variety of shoes crucial factors while buying, Are Discount and sales are correlated, Who gets influenced by advertisements and celebrities while purchasing footwear, Which factors in advertisements influence purchasing • The SPSS analysis gathered useful insights upon what attributes the consumers consider while purchasing by performing various parametric and non-parametric tests in SPSS & also suggested customization of shoes is preferred by youth.
Computer-Aided Decision Systems	**Designing a bank form in HTML and storing the data into MariaDB database** • Image of the webpage was given and from that, an HTML form was to be designed which contained number of fields to be filled by user • The values filled by user were passed to PHP script and through PHP the data was stored into database of MariaDB • Also, a webpage was designed for bank manager for viewing new users updated in database for approving the loan

COURSE WORK AND SKILLS

Relevant Courses	Data Mining and Knowledge Discovery	Probability& Statistics	Statistical Modelling for Business Analytics	Advanced Statistical Methods for Business Analytics	Computer Aided Decision Systems	Operations Research	Marketing Research	Introduction to Computing (JAVA)
Technical Skills	R(ggplot2, plotly, data.table, dplyr, rattle, rmarkdown, knitr, tidyverse, esquisse,MLR, Ranger, purr, parsnip, rtweet, Reticulate)	Java	MS Office (Excel, Word, PowerPoint)	Python (numpy, pandas, matplotlib, plotly)	R Shiny	PHP	SQL

ONLINE LEARNING & CERTIFICATIONS

- R Programming A-Z: R for Data Science with Real Exercises! (Udemy) & The Python Bible: by Zihad Yehia (Udemy)
- Certified Business Analytics Course: Linear & Logistic Regression, Decision Tree, Segmentation, Market Basket Analytics, Machine Learning (Eduonix)

POSITION OF RESPONSIBILITY

- Orientation Team Member (OTM) & Buddy, Counselling Service IITK

AWARDS AND ACHIEVEMENTS

- Secured All India score of 725 in GATE 2018, AIR 501 in UPSEE 2013