Ritima Malhotra, PhD

Data Scientist | Researcher | Educator

(647) 762 0643 | ritima1@gmail.com | linkedin.com/in/ritimam
ritima1.github.io | github.com/ritima1

Education

Post Graduate Program in Data Science and Business Analytics (PGP-DSBA) - University of Texas, Austin 2022

Doctor of Philosophy (Ph.D. in HRM/OB) - Punjabi University, India 2012-2018

Master of Commerce (e-Com) - UBS, Panjab University, India 2011

Bachelor of Commerce (Honours) - Panjab University, India 2009

Technical Skills

Machine Learning: Supervised and unsupervised Models - Classification, Linear & Logistic Regression, L1 & L2 Regularization, Clustering (K-Means, Hierarchical), Random Forest, Decision Tree, Bagging, Boosting

Statistical Methods: Predictive Analysis, Hypothesis Testing and Statistical Inference, a/b Testing, Principal Component Analysis and Dimensionality Reduction, Market Basket Analysis, Text Analytics

Tools & Languages: Python (NumPy, Pandas, Matplotlib, Statsmodels, Scikit-learn), SQL, MS Excel, SPSS

Data Visualization/BI: Seaborn, Plotly, Tableau, Power BI

Source Code: Github

Other Tools: SAFe Agilist, Waterfall, Agile

Professional Summary

A purpose-driven **data enthusiast** who takes pride in building models that translate data points into actionable business insights. **6+ years of progressive experience**, with a Post Graduate degree in Data Science and Business Analytics. Having conducted an extensive empirical study during PhD (used experimental research design to analyze data pertaining to emotional intelligence), I'm now eager to apply my data science and analytics knowledge to real-world business problems.

Projects

Project 1 : Trade & Ahead

- analyzed the stocks data to build a diversified portfolio using cluster analysis
- studied group characteristics via elbow plots & silhouette scores (K-means Clustering) and cophenetic correlation (Hierarchical Clustering)
- shared insights about given attributes to aid future investor decision making

Skills and Tools: Unsupervised Learning - EDA, Data Scaling, K-means Clustering, Hierarchical Clustering, Cluster Profiling, Dimensionality Reduction using PCA

Project 2 : ReneWind

- built various classification models & tuned them to find the best model that would help identify failures in the wind turbine generators
- aim was to repair the generators before they broke to reduce overall maintenance cost
- objective was to minimize the false negatives by maximizing the recall score (i.e., predicting generator failure correctly)

Skills and Tools: Model Tuning - EDA, Missing Value Imputation, Oversampling (using SMOTE), Undersampling (using RandomUnderSampler), Regularization, Hyperparameter tuning, Building Pipelines

Project 3 : EasyVisa

- built a predictive model (using classification) to facilitate the process of visa approvals
- identified factors that influenced case status to help recommend a suitable candidate profile that would help determine when to certify or deny visa to an applicant
- objective was to maximize F1 score to minimize both false negatives & false positives

Skills and Tools: Ensemble Techniques - EDA, Customer Profiling, One-hot Encoding, Split Testing, Decision Tree Model, Bagging Classifiers (Bagging and Random Forest), Boosting Classifier (AdaBoost, Gradient Boosting, XGBoost), Stacking Classifier, Hyperparameter Tuning using GridSearchCV

Work Experience

Partial Load Professor - Faculty of Business, Humber College (HR Analytics & Metrics, Applied HR Policies) Oct 2020 - Present

Part-Time Faculty - School of Business, Conestoga College Sept 2021 – Dec 2021

Data Analytics Instructor – Canadore College, Brampton Oct 2020 - Nov 2020

Data Analyst - Dawn2Dusk Irradiance Pvt. Ltd., India Sep 2017 - Sep 2019

Assistant Professor - Goswami Ganesh Dutta SD College, India Jul 2015 - Aug 2017

PhD Fellow - Department of Commerce, Punjabi University, India Sep 2012 - May 2018

Achievements

- Finished top of the class **(Rank 2)** in the PGP-DSBA course (October 2022)
- Awarded the **Python Fundamentals - Champion** badge in the PGP-DSBA course
- University Gold Medal holder **(Rank 1)** in Master of Commerce (e-Com), Panjab University, India
- Awarded **Junior Research Fellowship** by the University Grants Commission, India
- Emotional Intelligence Training Certification by Goleman EI (2020)

Project 4 : INN Hotels

- built a predictive model (using logistic regression and decision tree) to identify which booking is going to be canceled by customers in advance
- identified factors that had a high influence on booking cancellations to help formulate timely policies for cancellations and refunds
- objective was to maximize F1 score to minimize both false negatives & false positives

Skills and Tools: Supervised Learning (Classification) - EDA, Data Preprocessing, One-hot Encoding, Split Testing, Logistic Regression, Fixing Multicollinearity, Finding optimal threshold using AUC-ROC Curve & Precision-Recall Curve, Decision trees, GridSearchCV (for pre-pruning) & Cost Complexity Pruning (for post-pruning)

Project 5 : ReCell

- phone price prediction for used/refurbished phones was the objective of the project
- used multiple linear regression to build a dynamic pricing model (equation) and identified factors that significantly influenced the phone price (positively & negatively)
- all linear regression assumptions were tested and satisfied

Skills and Tools: Supervised Learning - EDA, Variable Binning, One-hot Encoding, Split Testing, Linear Regression, Linear Regression assumptions, Business insights and recommendations

Project 6 : E-news Express Project

- objective was to decide whether the new landing page of an online news portal (E-news Express) was effective enough to gather new subscribers or not
- control (old landing page served) and treatment (new land page served) groups were used
- variables like time spent on the new page, converted status, and preferred language were tested to conclude the effectiveness of the new landing page

Skills and Tools: Business Statistics - Hypothesis Testing (t-test, z-test, chi-square test, ANOVA), a/b testing, Data Visualization, Statistical Inference

Project 7 : Cardio Good Fitness Project

- objective was to identify customer preferences w.r.t three different treadmill models
- correlations via heatmap were studied to identify variable relationships that impacted customer buying decisions to target the right treadmill model to the right customer segment
- recommendations were provided to help Cardio Fitness increase their customer base

Skills and Tools: Python (Foundations) - Univariate Analysis, Bi-Variate Analysis, Multivariate Analysis, Missing Value Treatment, Variable Identification, Customer Segmentation, Correlation