Rahul Kumar

09899583701 / rahulin05@gmail.com

**PROFESSIONAL PROFILE**

A highly motivated and ambitious Machine Learning Engineer with 5 years of experience in the use of ML techniques to address complex real-world business problems. Experience includes predictive modelling, statistical machine learning, exploratory data analysis and natural language processing. Stays abreast of current and emerging trends, is passionate about predictive technology, and collaborates effectively in cross-functional team settings to translate business needs into technical solutions.

**KEY SKILLS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| * Applied Machine Learning | * Natural Language Processing | * Data Mining |  |  |  |
| * Pattern Recognition | * Online Learning | * Information Retrieval |  |  |  |
| * Text Analytics | * Image Processing | * Predictive Modelling |  |  |  |
| * Exploratory Data Analysis | * Model Evaluation and Selection | * Data Visualization |  |  |  |

**TECHNOLOGIES**

* **Predictive Modeling:** K-Nearest Neighbour, Random Forests, Decision Tree Learning, Ordinary Least Squares, Generalized Linear Model, Logistic Regression, Naive Bayes, Classification and Regression Trees
* **Statistical** **Machine Learning:** Regression Analysis, Correlation, Hypothesis Test, Pruning, Latent Discriminant Analysis, Principal Component Analysis, Bias-Variance Trade-Off, Regularization, Cluster Analysis, Ensemble Learning, Time-Series Analysis, and Apriori
* **Model Evaluation and Selection:** Area under ROC, Root Mean Squared Error, Cross Validation, Log Loss, Confusion Matrix, R-Squared
* **Exploratory Data Analysis:** Factor Analysis, Univariate and Bivariate Analysis, Imputation, Outliers treatment, Feature Extraction and Scaling
* **Languages and Others:** Python, C++, Web Scraping, scikit-learn, NLTK, Pandas, Syntaxnet, BeautifulSOUP, Bottle and SQL

**WORK EXPERIENCE**

**Data Scientist (B1)** at **EXL Services, Gurgaon Mar 2018 – present**

* Data Modelling and analysis

**Description**: To improve the response rate of an e-mail marketing campaign of a US based insurance client.

**Responsibilities and Achievements**: Worked on extremely skewed data for ~1 million customers with an event rate of approximately 0.05%; ***aim***: identifying customers likely to respond to marketing e-mails using information about them contained in 1100 variables. Achieved an F1-score of ~0.6 by fitting gradient boosted trees on the dataset, which was class balanced using a mixture of under sampling and Synthetic Minority Over-sampling Techniques (SMOTE)

* Email Classification for automatic labelling

**Description**: Text Classification system for multinational electronics manufacturer email system which classify each email ingested automatically into specified set of labels for inbox management.

**Responsibilities and Achievements**: Develop text classification pipeline for email classification using Hashing Vectorization and Gradient Boosting along with preprocessing using TF-IDF algorithm. Improved existing manual system and achieved overall 0.91 F1 score and reduced misclassification rate by custom analyzer.

* Document Classification and Entity extraction on legal lawsuits

**Description**: Classifying type of legal lawsuit document and extract relevant entity information from each document respective of its type.

**Responsibilities and Achievements**: Perform **Feature Engineering** on each document, develop text processing and robust machine learning **pipeline**, achieved F1 score of 0.82 with reduced misclassification rate with custom analyzer.

**Machine Learning Engineer** at **Genpact India Pvt Ltd** (Noida) **April 2015 – Feb 2018**

* People First Helpdesk

**Description**: Intelligent Helpdesk **Chat bot** with automated assistance for employees intended to provide information and support related to Genpact services and policies.

**Responsibilities and Achievements**: Improve Query Understanding with **Incremental Learning** and develop predictive models for **Named Entity Recognition** and Intent Classification, improved product performance by **online learning**.

**Technologies**: Python, Scikit-Learn, TensorFlow, Pandas, Bottle, Memorization, Gensim, SyntaxNet, Piwik, MatplotLib, Seaborn, JIRA, Pylint

* Customer Lead Scoring

**Description**: Predictive lead scoring for an American Multinational Client which access customer data from SAP (on-premise) Hana DB and predicts whether user will provide a lead in the future or not from user details and activities reported in database using Machine Intelligence.

**Responsibilities and Achievements**: Develop and evaluate machine learning pipeline for data ingestion and predicts lead score for each customer, **improves Business model** and customer engagement.

**Technologies**: Python, Scikit-Learn, HDBCLI, Pandas, Memcache, Piwik, NLTK, Bottle, Matplotlib, Seaborn, Jupyter, JIRA, Pylint

* Conversational AI for Genpact Neural Intelligence Platform

**Description**: Conversational AI is an Artificial Intelligence technology introduced by Genpact Neural Intelligence Platform for Natural Language Understanding and dialogue management.

**Responsibilities and Achievements**: Perform **Feature Engineering** for Semantic Annotated dataset, text processing and develop robust machine learning **pipeline**, increase classifier **F1 score** to **83** from **79**.

**Technologies**: Python, Numpy, Scikit-Learn, Pandas, SQL-Lite, Memorization, Bottle, NLTK, Gensim, MatplotLib, Seaborn, JIRA, Pylint

**Business Analyst at American Express** (**Gurgaon)**Oct 2013 to April 2015

* Worked on predictive **analytics problems** by surfacing right data, investigating abnormal trends and providing actionable recommendations to quantify the impact.
* Developed a **price prediction model** based on historic prices and macro-economic factors for a manufacturing firm to help them with procurement planning.
* Developed a **text analytics** model using call center rep's comments to predict which service issues can be resolved over phone, thus reducing agent visits, for a leading telecom provider in US
* Able to construct an analysis around a specific business question, write up a clear, useable project plan, and execute with minimal guidance and demonstrate compliance with internal procedures, processes and values.
* Collaborate with **cross-functional** groups such as Engineering, Sales and Legal to **increase automation**, update policies, fix loopholes and provide a better experience for consumers

**EDUCATION**

* **Bachelor of Technology in Electronics & Instrumentation, 2009**

*Uttar Pradesh Technical University, India*

**AWARDS AND RECOGNITION**

* Disruptive Thinker’s Award by CTO, Lean Digital, Genpact – 2016
* Bronze Award for active contribution in Genpact Neural Intelligence Platform - 2017

**CERTIFICATIONS**

* Machine Learning by Stanford University, Dec’16 – Mar’17 (4 months)