K Pradeep Kumar

8977425239 | pradeepkumar89774@gmail.com

Professional Summary:

- Dynamic and motivated professional having 4.5+ years of experience in IT industry which includes 3 years of experience in Machine Learning. Experienced in creating data regression models using predictive data modeling, and analyzing data mining algorithms to deliver insights and Implement action-oriented solutions to complex business problems.
- Having a thorough knowledge of mathematical and statistical concepts behind various supervised and unsupervised Machine Learning algorithms, analyzing large, complex, multidimensional data sets, developing analytical solutions and predictive analytics using Python.
- Exceptional analytical, interpersonal skills with proficiency at grasping new technical concepts quickly and utilize the same in productive manner.

Technical Skills:

Machine Learning: Linear Regression, Logistic Regression, Decision Trees, Gradient Boosting, Random Forest, Support Vector Machines, Naive Bayes & K-Nearest Neighbors, K- Means Clustering, Principal Component Analysis, NLP.

Deep Learning: Convolution Neural Network, Recurrent Neural Network, LSTM

Statistics: Hypothesis Testing, Z-Test, Chi-Square, ANOVA, Normal/Gaussian Distribution, Bayes theorem.

Python Libraries: Pandas, Numpy, Scipy, Matplotlib, Seaborn, Sklearn, NLTK, Keras, Tensorflow.

Data Base: SQL.

Professional Experience:

Working as **Data Analyst** for **Mindtree** from 01.04.2019 to till date
Worked as **Data Associate** for **Mindtree** from 10.11.2016 to 31.03.2019

Projects Summary:

Project 1: Demand Forecasting
Project Description: Build a model to predict the sales of the product given the market indicators and other factors that affect the demand and price of the product.
Roles and Responsibilities:

- Involved in understanding the data flow and business requirements.

- Selecting features, building and optimizing classifiers using machine learning techniques
- Identifying the missing values and performed missing values treatment using mean/median for the continuous variables and mode for the categorical variables.
- Performed Principal Component Analysis (PCA) for dimensionality reduction.
- Performed outlier detection using boxplot using the seaborn library in Python. Scaled the continuous variables using Standard Scaler and performed One Hot encoding for the categorical variables using the sklearn library in python.
- Checked for the correlation using Pearson Correlation & ANOVA techniques.
- Trained Linear Regression, Random Forest, Gradient Boosting models on the data and finalized with the Gradient boost model.
- Used the GridSearchCV & RandomisedSearchCV techniques in Python to perform hyperparameter tuning.

Project 2: Credit Card Approval.
Project Description: The goal of this project to predict the bank is able to decide to issue a credit card to the applicant or not.
Roles and Responsibilities:

- Understanding the business case and getting insights.
- Carried out Exploratory data analysis.
- Treated missing values in the data set with Mean/Median for numerical Features and with Mode for Categorical Features.
- Performed Scaling for numerical features & One-Hot encoding for categorical Features.
- Performed Correlation check by using the statistical methods Chi-Square & ANOVA.
- Trained Logistic Regression, Naive Bayes Classifier, models on the data and finalized with the Naive Bayes classifier model.
- Computed the Precision, Recall and f1-score using the actuals and predictions. Simulated the AUC curve using Specificity(or)True Positive Rate & False Positive rate.

Project 3: Sentiment Analysis
Project Description: Sentiment Analysis of customer reviews and feedback of customers about different products. We had cleaned the data to remove non-essential words, compared data for positive and negative polarities and provided a final score.
Roles and Responsibilities:

- Built Python Script to mine the unstructured text / data (by using NLTK framework)
- Performed the preprocessing steps like Tokenizing, Stopwords removal, Lemmatization, stemming to find the root words.
- Creating Bag-Of-Words model and TF-IDF and created a document term matrix.
- Identifying the clusters from unstructured data using K-means clustering.
- Trained with Logistic Regression and Navie Bayes MultinomialNB models and finalized with MultinomialNB.

- Tools & Techniques: Python, NLTK, Sentiment classification.

Project 4: Customer Medical Insurance Claim Prediction

Project Description: The aim was to develop a risk predict model based on the demographic information and pre-existing diseases given by the customer to decide the premium and design policies accordingly.

Roles and Responsibilities:

- Analyzed the modelling data using data mining techniques, including dimensionality reduction, data enhancement and data transformations.
- Managed imbalanced data by using imbalanced techniques like oversampling and under sampling and SMOTE.
- Performed Predictive Modelling for predicting claim of customer with different demographics and pre-existing conditions.
- Build a best fit model using machine learning algorithms (Logistic Regression, XGBoost, Random Forest).
- Evaluated the model performance using confusion matrix, precision, recall, f1 score & AUC-ROC Curve.

Educational Qualifications:

- Completed B Tech (Electronics and Communication Engineering) from Meenakshi College Of Engineering, ANNA University in the year 2012.