KARUNYA HALOI

Bangalore
Phone: 9071073604
Email: karunyahaloi01@gmail.com

OBJECTIVE

Gold Medalist from Bangalore University, is seeking a full-time position in an organization, where I will get the opportunity to work under the guidance of top- notch professionals enhancing my knowledge and allowing me to come up with tangible results

SKILLS

- Machine Learning
- Proficient in Python
- Time Series Forecasting
- Visualization
- Statistics
- Lidar, QGIS & ArcGIS, Terra maps creation

EDUCATION

- Gold medalist in Master of Science (Geography) from Bangalore university for the year 2015-2017.
- Scored 1st Class Degree in B.A (Geography) scoring 71.3% from Cotton University for the year 2012-15.
- Scored 87.2% in Class 12 from CBSE board for the year 2012.
- Scored 8.2 CGPA in Class 10 from CBSE for the year of 2010

CERTIFICATIONS

- IABAC CERTIFICATION FROM DATAMITES(CERTIFIED DATA SCIENCE)
- DATACAMP – PYTHON FOR DATA SCIENCE
- CODEACDEMY – LEARN PYTHON 3
- UDEMY- SUPPORT VECTOR MACHINES IN PYTHON(SVM-Python 2019)
- GUVI- MACHINE LEARNING 101
- GIS Intermediate Certification on QGIS 2.10.1, QGIS 2.18 and ArcGIS.

EXPERIENCE

NAKSHA TECH PRIVATE LIMITED
LIDAR SOFTWARE ENGINEER / Duration :1.9 years

Project: Pythonic **Data Cleaning with Pandas and NumPy** for Geospatial DATA obtained by LIDAR IMAGING for **Powerline Classification and Ground**.

PROJECTS

Title: TELECOM CHURN

TASK: TO FIND INFLUENCING FEATURES TO CALCULATE CHURN RISK SCORE for Predicting customer retainment

Libraries used: scikit-learn(sklearn),matplotlib, pandas,numpy,seaborn,sqlalchemy

Statistical techniques used: graphical representations, random sampling classify, correlation matrix, Smote, cross-validation

Approach Used: Checking the categorical and numerical features for making dummy variables, label encoding of specific feature columns and further dropping the unnecessary columns. Interpreting correlation and visualizing through bar graph, pair plot, and heatmap. Getting predicted probability for the target variable (1 & 0). Used **Random Forest Classifier** to get the accuracy score, crosstab, confusion matrix, precision score, and recall for the trained model. Checked the imbalanced dataset and used SMOTE to get better accuracy.

RESULTS: The final prediction can be visualized through a graph diagram to Predict the customer retainment.

CODE::https://github.com/Karunya15/KarunyaHaloi/blob/master/TELECOM_CHURN.ipynb

Title**:PIMA INDIANS DIABETES DATABASE**

TASK: The objective of the dataset is to diagnostically predict whether or not a patient has diabetes,

Libraries used: scikit-learn (sklearn), matplotlib, pandas, NumPy, seaborn

Statistical techniques: Logistic regression, graphical representations, correlation matrix

Approach Used: Used logistic regression ML model and predicted accuracy, split the dataset using train_test split(test_size: 0.1), viewed the classification using confusion_matrix , crosstab, and heatmap.

RESULTS: The Outcome can be observed in a graphical representation.

CODE: :https://github.com/Karunya15/KarunyaHaloi/blob/master/kernel397c2e45f6.ipynb

Title: **RESTAURANT REVENUE PREDICTION**

TASK: Predict the annual restaurant sales in different regional locations.

Libraries used: scikit-learn(sklearn),matplotlib, pandas,numpy,seaborn, datetime module

Statistical techniques: Linear Regression, graphical representations, correlation matrix

Approach Used: Used linear regression ML Model for predicting r2 score, split the dataset using train_test split, further created dummy variables, and performed label encoding for string values. Further correlating the data using a heatmap, scatter plot to show the city's revenue distribution and horizontal bar graph to show revenue.

RESULTS: The Outcome can be observed in a graphical representation.

CODE: https://github.com/Karunya15/KarunyaHaloi/blob/master/restaurants.ipynb

Title**: TITANIC SURVIVAL PREDICTION**

TASK: to create a model that predicts which passengers survived the Titanic shipwreck.

Libraries used: scikit-learn(sklearn),matplotlib, pandas,numpy,seaborn
Statistical techniques: Logistic regression, graphical representations, correlation, matrix, Smote
Approach used : Used the bar graphs and heatmap to study the features, worked on missing values, and labelled encoding for string values. Using the Logistic regression ML model to predict accuracy score. Further used train test split and used SMOTE for the imbalanced dataset to get better accuracy. Lastly testing the trained model on the test dataset.
CODE: https://github.com/Karunya15/KarunyaHaloi/blob/master/TITANIC.ipynb

RESEARCH PAPERS & PRESENTATIONS

- Presented findings on "**Impact of Climate Change on tea production in Assam**" in FIFTH INTERNATIONAL CONFERENCE ON GEOSPATIAL TECHNOLOGIES FOR URBAN FOREST AND CLIMATE CHANGE-PATHWAY TO SUSTAINABLE DEVELOPMENT conducted by UGIT
- Research Paper Published for Dissertation on "**Infrastructure and Economic Development of Kengeri ward – 159**" under the Department of Geography, Bangalore University.

All the information Provided are correct and best of my knowledge . Thanks for your time.

Karunya haloi
10th May 2020.