|  |  |
| --- | --- |
| **Module Code** | **CN7022** |
| **Module Title** | **Big Data Analytics** |
| **LSBF Coursework Instructor** | **Dr Preethi Kesavan** |
| **Coursework Title** | **Big Data Development Project (3000 words)** |
| **Weighting** | 100% |
| **Handout Date** | **October 2022** |
| **Coursework Submission Date** | **December 2022** |
| **Students’ Name & Students’ ID** | |  |  | | --- | --- | | **Student Name(s)** | **Student ID Number(s)** | | RAJAKUMARAN NISHANTHI | S1030554 | | VENKATARAMAN RAMU | S1030784 | | SONG JINGJING | S1031028 | |  |  | |
| **Documents uploaded in LSBF Canvas:** | ⭘ Assignment Report  ⭘ Presentation Report  ⭘ Source Code, Scripts & Output  ⭘ Turnitin Report |
| **UEL Turnitin Submission Similarity Index %** |  |
| Text, letter  Description automatically generated | |

I. Table of Contents I

**1.Executive Summary 2**

1.1 Introduction 2

**2.Project Background2**

2.1 Purpose and Scope2

2.2 Objective and Goals

II. Main Contents

**4.A Data Collection and Processing3**

4.1 Data Source3

4.2 Libraries 3

4.3 Data Demonstration/Exploration3

4.4 Data Cleaning and Imputation3

**5.** **Understanding of Hadoop and spark engine** 4

**6.Interpreting of Data /Data Analysis 10**

6.1 Model Building8

6.1.1 RandomForestRegression9

6.1.2 Linear Regression9

6.1.3 Gradient Booting Regression9

6.1.4 DecisionTreeRegression9

**7.Data Calibration (Measure Accuracy, Performance Metrics)** 4

7.1 Train and Testing Data 8

7.1.1 RandomForestRegression Accuracy and Metrics9

7.1.2 Linear Regression Accuracy and Metrics 9

7.1.3 Gradient Booting Regression Accuracy and Metrics9

7.1.4 DecisionTreeRegression Accuracy and Metrics

**8.Model Visualization 2**

8.1 Linear Regression Prediction with Visualization8

8.2 Random Forest Regression Prediction with Visualization8

8.3 Gradient Boosting Regression Prediction with Visualization8

**9.Comparative Analysis /Model Evaluation 2**

9.1 Decisions and Alternatives 8

**10.** **Testing and Predicting 2**

10.1 Testing Results with Dummy Values8

**11.Critical Discussion with Spark Engine and Other Relevant Analysis 8**

11.1 Model Limitations 8

11.2 Recommendations9

**12.Implications of Professional Practice and Future Research Development 14**

**13. Conclusion 13**

III. Table of Contents II

**14.References14**

**15.Appendices/Annex15**

15.1 Team member’s Contribution and Tabulated Meeting Log14

15.2 Gantt Charts/Time Plan

**1.Executive Summary**