**Section 1 : Topic Submission Form**

This form should be submitted by the mentioned deadline.

Name: Shushank Shekhar

Student Number: 1089169

Course: LJMU MS Data Science C17

**Fill your topic/s below**

Project Title/Area 1: Predictive analysis of retail sales

Dataset: <https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/poundsdatatotalretailsales>

Description: This predictive technique will incorporate Alternating Series Selection techniques using facebook prophet, which are based on the observation that sales data often consists of both regular patterns and irregular variations. By systematically identifying and extracting alternating series patterns, the research aims to refine the forecasting accuracy further. This hybrid approach will enable retailers to capture both the long-term trends and short-term fluctuations that impact their sales, leading to more precise forecasts.

The proposed thesis will involve collecting and analyzing historical sales data from a diverse set of retail organizations. This dataset will be used to train and validate the Prophet model, as well as evaluate the performance of the Alternating Series Selection techniques. The research will focus on identifying the optimal parameters and configurations to maximize the forecasting accuracy, considering factors such as seasonality, promotional activities, and external events like holidays or lockdowns.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Title/Area 2: Dynamic Modeling of Public Transportation Demand forecasting

Dataset: <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Description: The dynamic modeling of public transportation demand is an essential research area within the domain of time series forecasting. As public transportation systems face evolving patterns of ridership and demand, there is a need for advanced forecasting models that can adapt to these changes in real-time. This research aims to develop dynamic time series models specifically tailored for public transportation demand forecasting. The primary focus of this research is to investigate and propose techniques that enable the modeling of dynamic demand patterns in public transportation systems. Traditional forecasting models often assume stationary or slowly changing demand patterns, which may lead to suboptimal predictions when faced with sudden shifts, events, or changing travel behavior.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Title/Area 3: Forecasting Energy Demand in households

Dataset: <https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption>

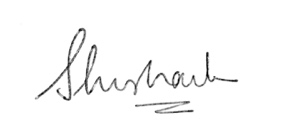
Description: The aim of this thesis is to investigate and enhance the accuracy and scalability of energy demand forecasting using the Prophet model. Energy demand forecasting plays a crucial role in energy resource planning, grid management, and ensuring efficient utilization of energy resources. The Prophet model, developed by Facebook's Core Data Science team, has gained popularity for its ability to handle time series data with inherent patterns and multiple sources of uncertainty. In this thesis, the focus will be on applying the Prophet model to forecast energy demand at different granularities, such as hourly, daily, or monthly intervals, depending on the dataset feasibility.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Fill in this section if a member of staff has agreed to be your supervisor:**

Member of Staff:                                                                                   \_\_\_\_

If you have found a supervisor then you and the member of staff who agreed to supervise your project should sign below.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_                                                                         Darshana Desai

Student Signature                                                                         Supervisor Signature

04-July-2023.                                                                    \_\_\_\_\_\_\_\_\_\_\_\_

Date                                                                                               Date

**Section 2 : Topic Selection Research**

**Table 1 : Topic 1 : Predictive analysis of retail sales**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Link to the Paper** | **Understanding of the Dataset** | **Understanding the Methodology Used** | **Dataset Link** |
| Short-Term Demand Forecast of E-Commerce Platform Based on ConvLSTM Network | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9303113/> | This article uses the real sales data set of Q company | This paper uses Long Short-Term Memory network (LSTM) and Univariate methods to predict sudden variations in time series |  |
| Predicting changing pattern: building model for consumer decision making in digital market | <https://www.emerald.com/insight/content/doi/10.1108/JEIM-01-2018-0003/full/html> |  | This paper proposes hierarchy models to forecast time series. Also intoduces fuzzy delphi logic. |  |
| Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data | <https://aircconline.com/ijcsit/V12N2/12220ijcsit03.pdf> | The paper uses real-world sales forecasting benchmark data obtained experimentally in a production environment in one of the biggest retail companies in Bosnia and Herzegovina | The framework is based on Facebook's Prophet algorithm. Paper also talks about using backtesting strategy to improve forecasts. |  |
| Retail forecasting: Research and practice | <https://www.sciencedirect.com/science/article/abs/pii/S016920701930192X?via%3Dihub> |  | Data pooling as a strategy to improve forecasts at lower product hierarchies. Examination of various characteristics and influential factors affecting product level retail sales.modelled challenges posed by marketing mix, promotions, and online review information |  |
| Comparative Study on retail sales forecasting between single and combination methods | <https://journals.vilniustech.lt/index.php/JBEM/article/view/1265/995> | The paper uses weekly sales data of an international furniture company operating in Turkey's retail sector for many years. | State space models (also known as ETS models). ARIMA (Autoregressive Integrated Moving Average) models. ANNs, Adaptive Network-based Fuzzy Inference System (ANFIS) |  |
|  |  |  |  |  |

**Table 2 : Topic 2 : Dynamic Modeling of Public Transportation Demand forecasting**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Link to the Paper** | **Understanding of the Dataset** | **Understanding the Methodology Used** | **Dataset Link** |
| Data-Driven Real-Time Online Taxi-Hailing Demand Forecasting Based on Machine Learning Method | <https://www.researchgate.net/publication/345494637_Data-Driven_Real-Time_Online_Taxi-Hailing_Demand_Forecasting_Based_on_Machine_Learning_Method> | The paper uses data related to the demand for online taxi-hailing services. | This paper uses Backpropagation neural network (BPNN) and then uses Extreme gradient boosting (XGB) to fine tune the model. |  |
| Predicting demand for air taxi urban aviation services using machine learning algorithms | <https://www.sciencedirect.com/science/article/pii/S0969699721000260> | The data used in this paper is related to predicting the demand for air taxi urban air mobility services during different times of the day in various geographic regions of New York City. | This paper uses algorithms like Logistic regression, Artificial neural networks, Random forests, Gradient boosting |  |
| Ticket Sales Prediction and Dynamic Pricing Strategies in Public Transport | <https://www.mdpi.com/2504-2289/4/4/36> | The paper has used a set of 3.23 million user-generated event logs of a bus ticketing platform to discover the factors that influence travelers in booking and purchasing bus tickets. | This paper proposes machine learning models for ticket purchase prediction and DA4PT methodology for discovering factors influencing ticket sales |  |
| Forecasting Taxi Demands Using Generative Adversarial Networks with Multi-Source Data | <https://repository.library.northeastern.edu/files/neu:m046qf889/fulltext.pdf> | The paper has used multi-source data including GPS taxi data, road network data, weather data, and points of interest for taxi demand forecasting. | This paper uses deep learning based model for taxi demand forecasting and also takes advantages of Generative Adversarial Networks (GAN) for fine-grained forecasting. |  |
| Taxi Demand Prediction Based on a Combination Forecasting Model in Hotspots | <https://www.researchgate.net/publication/343134819_Taxi_Demand_Prediction_Based_on_a_Combination_Forecasting_Model_in_Hotspots> | The paper has used GPS data from the Xi'an Taxi Management Office, which consists of vehicle location data recorded every 5 seconds for 30 days. The dataset consists of 40 million track points. | This paper uses baseline algorithms like Random Forest Model (RFM) and Ridge Regression Model (RRM) |  |
|  |  |  |  |  |

**Table 3 : Topic 3 : Forecasting Energy Demand in households**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Link to the Paper** | **Understanding of the Dataset** | **Understanding the Methodology Used** | **Dataset Link** |
| Deep learning for time series forecasting: The Electric load case | <https://arxiv.org/pdf/1907.09207.pdf> | IHEPC and GEFCom2014 datasets | This paper makes use of Feedforward and recurrent neural networks. This also proposes Sequence to sequence models and temporal convolutional neural networks. | <https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption> |
| Modeling and Forecasting of Energy Demands for Household Applications | <https://www.researchgate.net/publication/337009820_Modeling_and_Forecasting_of_Energy_Demands_for_Household_Applications> | The data used in this paper is the total number of houses for the years 1991, 2001, and 2011 in the districts of Brunei-Muara, Kuala-Belait, Tutong, and Temburong. This data was collected from the Government of Brunei census book. | Autoregressive integrated moving average (ARIMA) model and nonlinear autoregressive (NAR) neural network to forecast the number of houses |  |
| Forecasting of Energy Demands for Smart Home Applications | <https://www.mdpi.com/1996-1073/14/4/1045> |  | Uses Balanced deep learning architecture and Multi-objective optimization problem and fitness function |  |
| Uses Balanced deep learning architecture and Multi-objective optimization problem and fitness function | <https://link.springer.com/chapter/10.1007/978-3-030-64751-3_4> |  | Artificial neural network, Support vector machine, Gaussian-based regressions, Clustering |  |
| Data Modeling for Energy Forecasting Using Machine Learning | <https://link.springer.com/chapter/10.1007/978-981-16-0749-3_12> | The paper has used a dataset obtained from a commercial building to analyze the energy consumption pattern of buildings, specifically focusing on the heating, ventilation, and air conditioning (HVAC) system. | The algorithms from various families of machine learning like linear models, function-based learning models, lazy learning techniques, and tree-based techniques coefficient of determination. The interpretation of results obtained shows that K-nearest neighbor model from lazy learning family outperforms other regression models with reduced errors and improved accuracy. |  |
|  |  |  |  |  |