

Optimising the Mobile Cab Rental System App using Data Mining Techniques: SURVEY PAPER

S Swathi Rao¹, Dr S Krishna Mohan Rao²

¹Research Scholar, Mewar University

²Research Supervisor, Mewar University

Abstract- This study aims to explore and review the data mining and machine learning technologies adopted in research and industry to attempt to overcome the direct and indirect traffic issues on humanity and societies. This study has found that there is no standard traffic management approach that the community of traffic management has agreed on. This study is important to the traffic research communities, traffic software companies, and traffic government officials. It has a direct impact on drawing a clear path for new traffic management propositions. This study is one of the largest studies with respect to the size of its reviewed articles that were focused on data mining and machine learning. Additionally, this study will draw general attention to a new traffic management proposition approach. This study shows how mobile app is used to hire rental mobile can by using app. this is proposed by data mining and machine learning techniques.

Index Terms - Fuzzy rule-based classification systems; big data; map Reduce; Hadoop; Rules fusion, JUPITER SIMULATON TOOL

1.INTRODUCTION

The cab rental system has undergone a sea change in the last decade from the manual system to internet based online based system. The traditional cab rental service is highly manual and here the customer register for the cab by phone or come directly to the office, so it took a lot of time and resources and also related each process requires different resources causing the existing report data becomes difficult to manage. Then came the online cab rental system in which the customer place requisition for cab online and the travelling agent office arranges the cab and drivers and send to the customer place. The payment is directly sent to the cab rental offices. With the advent of GPS and increased Internet speeds, the mobile apps of car rental systems have emerged and completely replaced

the traditional manual and online rental systems. In mobile app cab rental system, companies like Uber and Ola, the customers and drivers are connected with the pre-determined fare and cab is reached within specified time. Provision for payment is done digitally and as well as through cash. The cab rental mobile app is user friendly and satisfying the customers. However, because of last minute booking cancellations of late the customer satisfaction service is getting affected. The problem is further aggrieved with the cab cancellations close to the trip start time, thereby causing passengers inconvenience. The problem could be tackled by accurately classifying the data of cab bookings using data mining techniques and design a predictive model there by forecasting the booking cancellations and take necessary steps so as to satisfy the customers. detection system

Background of the Study

Deep learning

Deep learning is a bit of an increasingly broad gathering of AI pro-cedars subject to fake neural frameworks with depiction learning. Learning can be overseen, semi-coordinated or independent. The proposed structure depicts sex area subject to Computer Vision and Machine Learning Approach using Convolutional Neural Network (CNN) which is used to remove distinctive facial com-ponent. In any case, the facial extraction is analysed, and best features are introduced which would be significant for getting ready and testing the dataset.

Deep neural network DNN

Profound neural systems (DNN) are intended to learn by Multi-association of layers to such an extent that each and every layer just gets the association from the past one what is more, gives associations just to the following layer in shrouded part. The info comprises

of the association of the information include space with the principal shrouded layer of the DNN. The info layer might be developed by means of feature detection or some other component extraction strategy. The yield stratum is equivalent to the quantity of classes for multi_class order or as it were one for parallel grouping.

Convolutional neural network

A Convolutional Neural Network (Convolutional net/CNN) is a Deep Learning calculation which can take in an information picture, allot significance (learnable loads and predispositions) to different viewpoints/questions in the picture and have the option to separate one from the other.

Potential cyber security threats:

Since cyber-attacks mainly come from malicious threats in communication networks, we review cyber-attacks in electric power systems, and provide an extensive analysis of network vulnerabilities under important use cases in the Smart Grid.

Artificial intelligence AI

The artificial intelligence AI perform various computer vision tasks and assistive technologies implementation for sentiment analysis, text mining, machine learning techniques, and deep learning techniques.

- The AI works on gender detection on social media using prediction and detection.
- Deep learning method, machine learning techniques ML.
- Assistive technology using on visually impaired, sensor networks.

Data mining techniques.

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs.:

How mobile app perform for booking rental system.

- Segment your audience. ...
- Identify the key benefit for each group. ...
- Allocate quantitative data. ...

- Compare your quantitative and qualitative data. ...
- Apply your analysis to a campaign. ...
- Analyse the results.

DATA MAPPING:

In computing and data management, data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks, including: Data transformation or data mediation between a data source and a destination.

In order to figure out how the data needs to be formatted, or mapped, it is essential to build a data mapping document. The data mapping document must include specifically the source and target data mappings. It must also include the primary key of all tables in source system.

Data mapping is the process of matching fields from one database to another. It is the first step to facilitate data migration, data integration, and other data management tasks. Data now comes from many sources, and each source can define similar data points in different ways.

Research Gap and issues

The primary targets of the current investigation are as per the following,

1.The research problem is that Proposed middleware in constructing realistic Smart Grid scenarios and conducting realistic cyber security experiments. The proposed model will demonstrate how a predictive model can be used to solve a business problem. The predictive model (machine learning model) for classifying new bookings as to find whether they will eventually get cancelled due to car unavailability. With the background covered in introduction, the plan is to obtain the Kaggle_YourCabs_training.csv file and partitioned it 70% for training and 30% for validation. For classification tasks, KNN, KKNN, Random Forest, Ada Boost, and Neural Network algorithms will be used; and generate error matrices, risk charts, lift charts, ROC curve, precision chart, sensitivity vs. specificity charts, and precision vs. recall charts for the validation data, and error matrices and ROC cures for the testing data for all algorithms used.

2.LITERATURE REVIEW

The information system designed to more closely manager's needs and the system set up as major computer application area. The Management Information System (MIS) as a computer-based system makes information available to users with similar needs [1]. Manager used the output information. The earlier studies shown that MIS could used to manage car rental, expected to accelerate as well as archiving services to customers better and safer, making it easier when required at any time [2] [3]. The online implementation of management information system provided and supported the customers for reservations, assist management in knowing rental car inventory at a specified time, to process transactions between branches car rental, transportation transaction processing, which supports satisfactory service to customers and support the company's operational processes [4]. Web based car rental information system increases the customers and help promotion [5]. To understand the patterns of the dataset, K-Nearest Neighbour algorithm is used for Classification and Regression. In this method, the input consists of the k closest training examples in the feature space [6]. Various Machine Learning Algorithms have been researched and employed in designing the predictive model [7]. The other machine learning algorithm employed is Artificial Neural Networks. ANN is a beautiful biologically inspired programming/computational paradigm enables a computer to learn from observational data. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output. ANNs are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found [8]

2.1 Aim & Objectives of the Study

Aim

The aim of this research work proposed on we the primary goal of proposed research work is development of a low cost predictive model (machine learning model) for classifying new bookings as to whether they will eventually get cancelled due to car unavailability. The predictive model should derive proper classifications (0=no cancellation or 1=cancellation). The proposed development has following subsidiary objectives: I. Study the source

dataset, detect anomalies, clean the data, and partition it into training/test data for data mining. II. Classify the dataset using a machine learning algorithm KNN and KKNN (Supervised learning). III. Employ Deep learning techniques (ANN) and train the Neural Network for straightforward application of optimization theory and statistical estimation.

OBJECTIVES

1. To understand the inner structure of these methodologies that have been proposed to overcome the imbalanced data problem in data mining.
2. To acknowledge the real challenges imposed by imbalanced datasets in Data mining techniques implemented.
3. To carry out a thorough discussion on the main issues to be addressed for future work on the topic.
4. To reduce the problem of network congestion problem for rental car booking system based on map reduce scheme.
5. To perform the bagging classifier method using ensemble methods for getting behaviour analytics.
6. To perform data mining, reduce algorithm using fuzzy based classification algorithms for execute consumer behaviour analytics along with Jupiter anaconda navigator tool.

3. RESEARCH METHODOLOGY

The big data term is used to describe the exponential data growth that has recently occurred and represents an immense challenge for traditional learning techniques. To deal with big data classification problems we propose the Chi-FRBCS-Big Data algorithm, a linguistic fuzzy rule-based classification system that uses the map Reduce framework to learn and fuse rule bases. It has been developed in two versions with different fusion processes. An experimental study is carried out and the results obtained show that the proposal is able to handle these problems providing competitive results.

- Volume: This characteristic refers to the huge amounts of data that need to be processed to obtain helpful information.

- Velocity: This property states that the data processing applications must be able to obtain results in a reasonable time.
- Variety: This feature indicates that the data can be presented in multiple formats; structured and unstructured, such as text, numerical data or multimedia among others.

3.1 Research Design

This learning will utilize and expressive examination strategy. The spellbinding examination strategy will be embraced to break down the accessibility and openness of information on interpersonal organizations. The fundamentally inquire about work structure along these lines for execute expectation of cyber-attacks detection using IOT applications. The design of a study defines the study type (e.g. descriptive, correlational, semi experimental, experimental, review, meta-analytic) and subtype (e.g., descriptive longitudinal case study), independent and dependent variables, research questions, hypotheses, experimental design, and if applicable, data collection methods and a statistical analysis plan.

3.2 Data Collection Strategy

Primary data: The various survey paper collected for getting revise of work related to smart grid computing. Secondary data: data set collected form kaggle.com and perform and analysis of various map reduce algorithms works based on data analysis.

Data set:

- Real data set for performing consumer behaviour analytics.
- Kaggle .com provides data set.

3.3 Pilot Study

Pilot study will be undertaken to validate the research instrument used for the data collection on social sites. For the pilot study link of the data storage and collection will be shared with various social platform like UCI REPOSITORY sites. Pilot study will help in improvising the implementation work and further assist in calculating sample size for the study.

3.4 Sampling

In the survey will be distributed to respondents in the selected areas and also through various resources. Moreover, participants will be further asked to forward the link and draft to other contacts to facilitate

data collection. The sampling is depend on collection of various parameters which is depend on social networks data storage.

3.5 Sample Size

Sample size for this study will be based on web data like CSV file data storage. More than 1000 records, 50 attributes, 2000-3000 instances.

3.6 simulation Tools Used

For Analysis, a few apparatuses will be utilized to break down the information gathered through inspecting. An enlightening measure like mean and standard deviation will be utilized for rundown of huge information. Various types of diagrams will be utilized for graphical portrayal of information. For identification and forecast the different equations are executed by utilizing characterization technique.

- Accuracy
 - Sensitivity
 - Specificity
 - Precision
-
- Simulation tool using for implementation of prediction cyber-attacks analysis is PY-charm tool and python language.
 - Methodology like CNN convolutional neural networks, DNN deep learning neural networks, classification techniques, and algorithms like classification algorithms also representation of prediction methods.

4. SIMULATION DETAILS

The research work details have performed on the basis of 12 module set which perform the simulation tool detail JUPITER anaconda navigator 6.3.3 and panda import libraries.

The simulation work detail done on the basis of module performed on python code and Jupiter anaconda navigator and perform libraries for credit fraud detection using tensor flow method.

The simulation tool details which define the following points are given below:

1. JUPITER ANACONDA NAVIGATOR simulation tool details.
2. Python code details.

3. Panda import libraries.
4. Performance matrices.
5. KAGGLE real data set.

5.EXPECTED OUTCOMES

We present the experimental study carried out using the Data mining algorithm over data problems. We provide some details of the classification problems chosen for the experiments and the configuration parameters for the methods analysed.

Regarding the parameters used in the experiments, these algorithms use:

- Three fuzzy labels for each attribute.
- The PCF to compute the rule weight.
- The winning rule is used as fuzzy reasoning method.

6. SCOPE OF THE STUDY

The scope of study is related to latest problem resolving of social networks. Scope: The proposed research scope is to develop a machine learning model that can be employed by cab companies to reduce cancellation costs and improve customer satisfaction. Limitations: The research is tightly coupled to the available dataset. Any change in the dataset's parameters or its cardinality will need a few minor changes in the model. De Limitations: The cancellations refer genuine reasons. The cancellations due to unforeseen conditions like calamities, floods, Earthquakes are not confined to this research.

7. CONCLUSION AND FUTURE WORK

This paper gives the approaches of mobile car rental issues for using mobile app. the data mining algorithms and machine learning new approach for optimize the overall investment and resolve the network issue. The paper also describes the GIS advancement of this rental car system monitoring and take proper feedback from this data mining technique.

REFERENCE

[1] R. Mcleod and Jr. G. P. Schell, Management Information System, Tenth Edition. India: Pearson Education, Inc, 2007.

[2] Y. Damayanti, "Perancangan Sistem Informasi Penyewaan Mobil Rama Rental Car Dengan Menggunakan Microsoft Visual Basic Versi 6.0," unpublished. Undergraduate Thesis. Jakarta: Gunadarma University, 2005.

[3] M. R.T. Cahyanto et al., "Sistem Informasi Manajemen Rental Mobil CV. Opal Transport dengan Java SE (Standart Edition) dan MySql Berbasis OOP (Object Oriented Programming)," unpublished. Undergraduate Thesis. Yogyakarta: STMIK AKAKOM, 2010.

[4] Yodiyanto, "Analisis dan Perancangan Sistem Informasi Rental Mobil dan Angkutan Travel Berbasis Web pada PT. Kembang 88," unpublished. Undergraduate Thesis. Jakarta: Bina Nusantara University, 2006.

[5] R. D. Sari, "Building Application System Car Rental Reservation and Payment Online Web-Based (Case Study in The Rental Daras Corporation)," unpublished. Undergraduate Thesis. Bandung: Unikom, 2011.

[6] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[7] <https://inclass.kaggle.com/c/predicting-cab-booking-cancellations>

[8] <http://neuralnetworksanddeeplearning.com/>.