



# **COLLEGE OF ENGINEERING, PUNE**

(An Autonomous Institute of Government of Maharashtra.)

**Department of Computer Engineering and Information Technology**

## *PROJECT ABSTRACTS*

**BTECH INFORMATION TECHNOLOGY 2015-16**

---

### **COMPUTATIONAL LEARNING ANALYTICS FOR AGRICULTURAL PRICE PREDICTION**

In this project we have proposed different algorithms to predict agricultural prices. Here we have first gathered the data from the web sites that have agricultural data over the past few years using data mining. Then we have filtered the data according to our needs. We have applied different algorithms to the data, studied the accuracy, time taken to compute the result. For producing the results we have taken into consideration some factors namely total area harvested in the country, transportation cost etc. which will affect the prices in the long run.

### **CLOUD BASED PLATFORM FOR INTERACTIVE 3D VIRTUAL WALKTHROUGH AND FLOOR PLAN**

This platform provides an end to end automated system for creation of 3D interactive Walkthrough just using the 2D Floor plan.

The idea behind the project is to solve the problem of creating an efficient and easy to use cloud based virtual walkthrough system with nearly real- time syncing for architecture and interior design visualization. This project builds a cloud-based platform which enables the user to build, modify and view a floor plan on mobile, desktop and VR headsets which will dynamically be converted to 3D Interactive Virtual Walkthrough using a node-based approach for procedural generation.

The project allows multiple users to edit and update a particular floor plan from different devices and hence it is important that the changes are being persisted across all devices using the floor plan. The virtual walkthrough system created using these technologies will not only have to be efficient but immersive and highly interactive as well. The system designed will link up various devices (clients) to a central database (server) which is connected via a cloud network. The client is responsible for most of the work, that is, rendering of 3D graphics, walkthrough, floor plan creation, etc. and the server's responsibility is to make sure all the clients are in sync - and provide database access.

## **ARTIST PORTAL**

The main objective of artist portal web application is to provide the user information about the various opportunities available in the field they are interested in and want to pursue it as a career. They will be able to upload their information and interest details according to that will be suggested the events to participate and details about events related to the field or the courses and institutes if needed the prior training.

This web application also gives opportunities to the provider that includes the institutes or academies providing classes or training in particular field to upload information about it and even event handlers of various art forms can use this platform to give information about the same.

### **H-Eat - Heat Mapping for Restaurants around Pune**

Internet usage has exploded in the past few years due to the rapid boom in web technologies and the amount of power that is packed into it. The growth of web technologies is revolutionizing the way we live by changing the way we carry out everyday tasks like ordering food or searching for a better restaurant. The stage we are at right now is however just a starting point that is going to lead to something much bigger and life changing. The reason web technology is so interesting because of its vast features like geo-tagging and heat mapping. The future of web technology thus depends on making effective use of these features.

**Heat maps** are better to understand and visualize data as compared to regular maps. Restaurants obsessive's want to know which is the most trending and favorite food joint at some place. Hence we would plot a Heat Map which will show continually where the food lovers are flocking at that moment even if that might be an old spot using data based on check-ins retrieved from Foursquare API.

## **CONFIGURABLE PAYROLL SYSTEM**

Paying employees has always been a crucial, fundamental part of any business. Everyone has to pay their employees. But what does one do when one has a wide variety of employees being paid a with a wide variety of pay structures? No single program seems to be able to handle all the different pay structures without having to undergo modification. The Payroll Project is designed to remedy this. SAP has for long been held as one of the best ERP systems for any organization. The ability of SAP software to handle huge volumes of data efficiently is well known. Coded in ABAP/4 (SAP's proprietary programming language), The Payroll Project is the payroll system that can handle any pay structure and any variations in pay structure with a few simple adjustments. The secret to it's exceptional versatility? It's design, which focuses on the program being driven by the data from the contents of a few configuration tables in a database rather than relying on hard coded data. As the contents of the table change, the behavior of the program changes.

Therein lies the distinct advantage of The Payroll Project over any others.

## **PERSONAL SHOPPING RECOMMENDATION**

As e-commerce is developing rapidly, it is becoming a research focus about how to capture or find customer's behavior patterns when they shop online. Recommendation systems have proved to be a useful mechanism to identify customers' online behavior patterns and help both the customer and the seller to increase the sales of any given product.

Our project is based on providing recommendations to customers shopping via any electronic commerce means. We have the recommendation process based on customer's location and age. For new customer the recommendation will be given based on location, age. But for the existing customers we use his/her shopping history for recommendation of products. Due to this system the frequency of buyers will increase and sales of goods will be improved.

## **MORPHOLOGICAL ANALYSER FOR PALI LANGUAGE**

Natural language processing is a technique that is focused on extracting meaning from human or natural languages for the computer to perform operations on it in order to improve the interaction between humans and computers. So far the computer scientists have made numerous approaches in this particular field and as a result of their success we are able to perform computational operations on natural languages. The efforts made for NLP of English has resulted in development of an entire system required to understand and derive meaning from human inputs by the computers. This has led to development of English WordNet, English parser, English Tagger and many more utilities required for computational linguistics. Similar efforts have been made for many European and Indo-Aryan languages. However, there are many such languages which are still far away from being able to be processed by computers. Pali language also known as language of Buddha is still, far away from being processed computationally. Pali is a native language of the Indian subcontinent and it hosts the extant literature of Buddhism. Due to development of many popular regional languages the use of Pali has decreased over time. This has led to difficulty in understanding the scriptures and preaching of Buddha to people not having thorough knowledge of Pali. Thus there is a need of digitalization of the Pali literature for the use by common people. This project is our sincere effort to understand the process of Natural language processing and start its implementation for Pali language. Since the project is first step towards the computational understanding of Pali, it may need to be revised and improved over the period of time.

## **HIVE OVER COUCHDB**

Apache CouchDB which is also called as CouchBD is an open source database software that focuses on ease of use by using document-oriented NOSQL database architecture. The applications interact with CouchDB via HTTP. The number of actions that we can perform on CouchDB data is very limited such as creating and deleting a database, inserting and retrieving

data from database. These limited number of functionalities make the use of CouchDB less desirable.

Apache Hive is a data warehouse infrastructure built on top of Hadoop. Hive supports analysis of large datasets stored in Hadoop's HDFS. It provides an SQL-like language called HiveQL with schema on read and transparently converts queries to map/reduce.

By considering the limitations of accessing data stored in CouchDB and the functionalities that Hive provides, our project aims at accessing CouchDB data through Hive so that we can perform all SQL operations on that data.

## **SECURE DATA FORWARDING IN DISTRIBUTED DATA ENVIRONMENT**

When it comes to defend the system against the attacks on confidentiality and integrity, then authentication is the starting step. Traditional password based systems are very simple to implement but they have to face different attacks. As a repository to this, the token substituting and biometric based authentication systems were introduced further. However, these systems have not been improved to that extent that can justify the speculation. Thus, it is developed as an alternative to the password based login methods. But it also suffered due to the increasing screen dump attacks. We have designed structure of Graphical password Authentication system, which is not vulnerable by other authentication methods for common attacks. Encryption schemes protect data confidentiality. In the proposed system we have integrated a secure system with distributed storage that is formed by amalgamation of a scheme having threshold proxy reencryption with a decentralized ensure code. The distributed storage system supports robust and secure data storage and retrieval system, and also enables without retrieving the data, send data between any user.

## **STUDY OF CLUSTERING ANALYSIS USING NEWS DATASET**

"Study Of Clustering algorithms Using News Dataset" is a system for data analysis. We focus on different clustering algorithms on dataset to analyze which clustering algorithm is best for given data. The goal of this project is to analyze different clustering algorithms on given dataset and check which clustering is best and why. We performed document preprocessing on input datasets. Thus we focus on developing a system which processes the input dataset and return which distance measure technique and clustering algorithm is best for given dataset.

## **ADAPTIVE LEARNING SYSTEM FOR ANALYSIS OF PERFORMANCE AND LEARNING BEHAVIOUR OF SCHOOL CHILDREN**

In today's world, where information is vast and the journey to seek knowledge is extremely high paced and competitive, it has become imperative to identify the best learning methodologies. Our project comprises the use of psychological concepts like brain dominance and multiple intelligences along with continuous performance analysis to cater to each school student's

individual needs and make the learning process efficient and more enjoyable. The need to upgrade current teaching methodologies has resulted in several Intelligent Learners like Cognitive Tutor, Knewton Adaptive Learning, ALEKS, etc. These systems have been built with adaptive learning as the basis; however none of them fulfill the need of understanding each student's distinguished learning style. Our learner addresses the need for modifying teaching through a variety of means by using psychological concepts. Moreover, our learner caters to a different demographic of the student society. It uses real-life data for training and adaptive purposes. Our target age group is students who are 5 to 10 years old and currently studying in Government schools. The adaptive learner uses the concepts of machine learning and data analysis to incorporate the above mentioned features in the learning process of school students. As a whole, it fulfills the following objectives-

1. Intelligently identify the brain dominance using classification in machine learning.
2. Recommend efficient teaching techniques based on dominant intelligences.
3. Develop a learner that analyzes performance and modifies testing procedures accordingly, for maximum performance improvement.
4. Analyze performance measures like retention and grasping across the three brain dominances using graphical data analysis and regression in machine learning.

The learner implements the above mentioned objectives as follows-

1. It uses the Random Forest Classification Algorithm to classify a student into Left, Right or Middle brained and provides a unique teaching approach based on his/her dominant multiple intelligences.
2. It then uses an adaptive testing algorithm that continuously analyzes the student's performance and uses the inferences to adapt to his/her current skill-set.
3. The learner pays special attention to the following performance measures
  - a. Grasping: A linear regression model is used to compare the student's current performance to a standard grasping measure and adapt the system to this comparison's inference.
  - b. Retention: Test scores and time of exposure to each topic are used to draw exponential growth and decay curves. These help the learner to determine the time period after which a topic must be revised with the student.

Hence, the system is a holistic intelligent learner that takes into account the following factors-

- Psychological: Brain Dominance & Multiple Intelligences
- Past Performance
- Grasping Measure
- Retention Measure

These factors put together provide each student a unique learning experience.

## **VOICE ACTIVATED HOME AUTOMATION**

Home automation is a growing Industry. This is developed to provide help for the old and disable people. Controlling Home appliances is the working of Home Automation. In home automation, a wireless communication link must be provided to the remote user. Main task of this system is, make a system who will control household electronic things remotely via Voice Command. This project explains the design and structure of wireless home automation on the basis of voice command which has been built and implemented. This system controls all the electrical appliances through voice commands in a office or house.

Today there is a wide use of the computers and wireless communication generating large amount of data (big data). We generate and use such data. Data mining means extraction of information from data. As large amount of data is being continuously stored in database, we need to manage the database and retrieve the necessary information.

Data has a structure. We need theory to extract that structure from data for understanding the process and make the prediction for the further use in future. Data is cheap and available in large amount. So we are developing the tool for extracting the knowledge which is expensive and scarce from large data. In the proposed system we are trying to find out the distribution of datasets and accordingly system will help user to find out the strong classifier and weak classifier for particular dataset by using WEKA. We are implementing the Ensemble method to combine the performance of the strong classifier and weak classifier. The proposed system will help the user to find out more efficient classifier for that particular dataset.

## **PERSONALITY BASED USER CONTROLLED DYNAMIC RECOMMENDER SYSTEM**

The principle on which a recommender system works is that it suggests users the items they are likely to purchase. But, in most recommender systems, users are not provided explicit control over personalization. They are always left guessing on what basis they are being given recommendations. This is of great importance to recommender systems.

Moreover, unseen and novel recommendations ignite a user's interest in products and provides an engaging experience in viewing and buying products. As a result, recent recommender systems are concentrated greatly on providing adaptive novel recommendations which would be helpful for businesses. The main focus is building a recommender system and evaluating the same considering the above factors which will be scalable across all platforms.

## **APPLICATION OF BEACON TECHNOLOGY FOR RETAIL INDUSTRY AND ITS ANALYSIS FRAMEWORK**

This project describes the design and integration of the Bluetooth Low Energy (BLE) Beacon System in the Retail industry and its impact by analyzing the data generated by the system. When the exchange of signals take place between an android Smartphone and a beacon, the data

is generated and stored in the database, provided that the Smartphone and the database server are in the same network. The analysis of the data is done using few complex queries in the analyzing tool. Using this system, the shop owners can attract the customers roaming near the shop and they can further analyze the movements of customers for the benefit of the business.

### **MYLISPEAK ENGLISH LEARNING SYSTEM**

Importance of English Simulator Technology is a platform to develop your listening and speaking skills in English, improve communication skill and study the way to improve personality, and devise an intuitive way using audio video recording techniques ,videos, voice recognition techniques. And develop web based portal for students to improve their skill. The Project aims to improve the communication skills of the students and all, especially those who need to improve their language skill. Meanwhile, there are advances in Human computer interaction recently, with voice and handwriting technologies showing great promise. Technology like Espoir smart interviewer, Spoken English technology provide software based interactions for communication skill activities, whereas technologies such as BEC are helps to improve listening as well as reading and writing skill . The goal of the project is to improve communication and language skill and closely study the way how the professional person communicate. This could be through voice recognition technologies, recording and webcam technologies, a combination of both, or even some other third technologies.

### **AUTOMATIC QUIZ QUESTION REPHRASER**

With the expanding interest in quizzes and tests, there is a need to obtain an extensive question bank with diversity of questions. Today, such a task is performed manually and utilizes considerable amount of time and resources. In order to assist the quiz creators and the test takers, we propose an Automatic Quiz Question Rephraser which takes a question answer pair as the input data and modifies the input data to generate rephrases for that question as the output. We categorized the rephrases for a diverse set of questions. The fundamental categories include replacing with answer category in which the main noun phrase is replaced with the answer, the true-false category generates questions which will have True/False or Yes/No as the answer and the Substitution By Synonym category in which the verbs, nouns and adjective are replaced by their equivalent words. We carried out the parse tree analysis for the input question and rephrases to devise rules for generating rephrases belonging to different categories. Based on these rules, the system was programmed. Preliminary experimental work carried out on this system shows that approximately 75 percent of the rephrases are correct.

### **DESIGNING AND IMPLEMENTATION OF PAYOFF LANGUAGE FOR STRUCTURED FINANCIAL PRODUCTS**

"Designing and implementation of payoff language for structured financial products" provides a

platform/tool for those people who work in finance sector and wish to design a financial product, by giving them a simple, understandable language(YAML) with predefined set of data types that include finance jargon. Also, this project will help them to have exact visual representation of that product using YAML to XML Parser. XML to YAML parser in our project will give YAML code representation for the XML form of finance product.

Any product in Finance is well analyzed if its function is estimated, on basis of all given credentials in its design. Python representation of product will do the same job. YAML to python will give functioning python code for product that will tell us calculated payoff. Similarly, Python to YAML Parser in our project will give YAML code for any python input file of product.

In short, this project is a parser which will parse YAML to XML, XML to YAML and also YAML to Python, Python to YAML.

## **CAMERA BASED ATTENDANCE SYSTEM**

Camera based attendance system aims at automating the process and reduce time as much as possible. In current project we have tried to optimized the attendance system using face recognition, eigen functions, CxImages and visual studio. Most important part of whole project was learning and the experience of developing software product. This project can be further extend and can be connected with CCTV, smart phones, 3D camera, etc. Multiple face recognition can also be achieved with reference to our study.

## **DESIGNING A VISUAL PROGRAMMING LANGUAGE FOR DEFINING STRUCTURED FINANCIAL PRODUCT**

A structured product, also known as a market-linked investment, is a prepackaged investment strategy based on derivatives, such as a single security, a basket of securities, options, indices, commodities, debt issuance and/or foreign currencies, and to a lesser extent, swaps. These products are complex and difficult to understand for clients. Various financial communities such as banks, companies regularly develop new products for clients. Hence to invest in such products client first need to understand product and should be convince to use such products. Sometimes it becomes very difficult to explain functionality of such complex products to clients. A visual approach is best way to define a functionality of the complex products. Hence we come up with an idea which visually define functionality of complex structured product with drag and drop approach. As it will be efficient for both client and banks to create, identify and explore structured financial products.

Google Blockly is an open project of Google which deals with drag and drop approach to learn different coding languages. We use this Project of Google as a base and implement our idea into it.

With the help of our project user can be able to create structured financial products by simply

drag and drop of the various blocks which include basic finance blocks, data types, mathematical blocks etc. The created block can be saved for further references. By running the respective block it will be generating respective payoff of that product according to python code calculations.

It will also generate financial code which display result in simple financial language. The graphs are included for easy understanding of particular products. Some financial products are defined according to graph generated. Here user can import graph points of particular financial product and it will display chart graph. User can also be able to access database for saving clients profile or to save some payoffs.

## **SYNTACTIC AND SEMANTIC PLAGIARISM DETECTION BETWEEN DOCUMENT**

In this era of internet, there are lots of emerging technologies for accessing information. Learning has become very easy with lots of material to study. But at the same time, available resources are being exploited to easily copy and kidnap the contents or ideas of others and make them their own. Academic reports, various research documents are more prone to plagiarism. Because of the need to detect the textual kidnapping, many techniques and methodologies have been developed for detecting and preventing plagiarism thus maintaining originality of contents.

## **INTRANET BASED MESSAGING SERVICE**

The communication between employees in same office is immutable. For such communication we have resources which needs internet connection example WhatsApp. In our system the connection of internet is not required which will minimize the communication cost. This system is secure because no outsider can approach the system as he is not aware of password of Wi-Fi and also he has to be in the range of network. First of all, app will show all the connected/available devices in its range. All the devices which are connected to router are able to communicate with each other. We selected android as platform for our system as out of ten people nine uses android smartphone. Android is open source software.

## **VISITOR MANAGEMENT SYSTEM**

Daily, many people come to visit an organization or an event. Maintaining their record manually is the tedious job. Also information retrieval from this data takes a lot of time. We cannot perform any statistical analysis on that. Data received is not well formatted and tabulated which makes it difficult to find required data.

We are proposing to develop a system which provides an automated, simplified and generalized solution to this problem, by creating a web based application which uses database to store data which can be further queried to provide filters on visitor like gender wise visitors, frequent

peoples, rare people, visitors on particular duration (on particular date, time), details information about the visitors with some other additional features. Main objective of the system is to provide an organization a better insight of the visitors and show summarized data which will be helpful to manage their resources more efficiently.

### **COTTON PICKER MOBILE APPLICATION**

Indian economy is agriculture-based. Cotton is an important factor in agricultural industry. According to studies, there was 7.6 million hectares of land under cotton production with average yield of 307 kg/hectare in the year 2003-04. Though the number sounds big, a lot of cotton is wasted due to manual picking of cotton. Also finding the labor is difficult task. Hence, John Deere is trying to use cotton-picker machine.

Cotton-Picker is the machine which collects the cotton with an efficiency of almost 95 %. However the machine is costly and cannot be afforded by a single farmer. So, the machine is generally owned by the contractor. The contractor has multiple field with different geographical conditions. So to increase profit of contractor, it is necessary to analyze and then manage the available fields. This can be achieved through cotton-picker application.

The application takes the input as field and machine details and provides approximately estimated cotton collection, expenditure and profit that can be generated after harvesting the given field. There are three phases of this project:

Phase 1 : In this phase, we simply take the input and provide output. It is a simple single field analysis.

Phase 2-a : In this phase, we compare different fields and then provide the comparative analysis of the fields.

Phase 2-b : In this phase, we compare different fields on the basis of not only their details but also their positioning. As such it provides the optimised path between the fields so as to maximize the profit maximally.

### **REMOTE MOBILE CONTROLLER AND ANTI-THEFT APPLICATION (ANDRIOD APPLICATION)**

The use of smartphone is increasing exponentially with the rapid change of technology. However, as the use of smartphone has increased, smartphone theft have also increased. Today, smartphones have different apps that have us logged in. These apps contain our private data, and it becomes very important for us to secure this data in a situation when our smartphone is lost or stolen. This paper provides a model to protect the smartphone from the theft and also the security of the user. It will also help in accessing a smartphone with the help of any other remote phone (smartphone or any other simple mobile phone). It has functionalities that enable a remote phone to send commands to a smartphone. These commands then trigger different actions on the smartphone. The application will be used to retrieve data from the applications

installed on the smartphone to the simple phone. It will also enable to use the application over the internet. Also the model will help us to simulate a smartphone screen onto other smartphone.

## **INTELLIGENT TRAFFIC SIGNAL SYNCHRONIZATION USING FUZZY LOGIC AND Q-LEARNING**

In the past decade, due to industrialization, urban traffic has increased tremendously. Due to which, the urban population has to invest more amount of time in traveling. Increased road traffic results in an increased number of accidents and more consumption of fuel, thus wasting energy. Hence for solving this issue, a traffic signal synchronization system has been proposed which takes real time traffic signal data as input and with the implementation of multi-agent fuzzy logic, it introduces the design of an intelligent system which would smoothen the overall road traffic of the city. Fuzzy system can handle the various levels of uncertainties found in the input data taken from the traffic signals. Since fuzzy logic system needs expert knowledge for its rule base and the rule base remains unchanged once defined, this paper adds up *Q*-learning module so that the system learns by itself by updating the set of rule base.

## **STACKOVERFLOW DEVELOPERS AND TECHNOLOGY ANALYSIS**

For development of the any of the software, it is necessary to know which technologies should be used. But users generally don't know the all development technologies and features of the available technologies. So, from this project we will analyze technologies and we will explore more about their characteristics. The Stackoverflow gives us a technical forum for users to ask question and, through involvement on stack exchange site and users' participation, to either upvote or downvote questions and various answers, editing questions and answers. Stack Overflow is a social programmers website where people from various programming/development backgrounds post questions for the problem which they are facing and other people who know the solution for their problems will post the answer for those questions and if the owner of that question i.e. the person who post the question found the correct solution for his problem will upvote the solution.

Stack overflow provides not only the information about various technologies but also various problems and associated solution to the problems. We are able to use the data to analyze on various aspects. Stack exchange provides APIs for getting the information about the technologies. With the help of those information we can suggest users about the popularity of languages/technologies, problems in specific languages, domain in which improvements are needed for languages, technologies etc.

With the help of stackoverflow data we may get to know about developers who works in different technologies and we also get to know about their expertise in each of technology they are using. So we can advise the list of developers to whom our users can contact for their project development whenever needed.