**BONAFIDE CERTIFICATE**

Certified that this project report titled **Smart User Targeted Advertising** is the bona fide work of **Saransh Agarwal, Mayuri Rani, Palak Gupta** who carried out the research under my supervision. Certified further, that to that best of my knowledge that work reported here in does not form the part of any project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. Annapoorani K Dr. B. Amutha

(ASSOCIATE PROFESSOR) (HOD)

**ABSTRACT**

The factors that manipulates the prices of the product depends upon demand which in turn is basically dependent on a click rate and time together. But Other factors that defines todays advertisement include user’s preferences that is derived using machine learning and demand and market economy. But our study focuses on some new factors that is to let the publisher decide the price at which he wants to advertise the product. Another factor is orientation based suggestions that determines what kind of ad a person will se. This is an example where males are shown more male oriented ads and females are presented with more of female oriented ads.

In this method, the prices of product are regulated using time base and counter based. In time based the prices are increased as time passes and is also based on number of clicks. These both factors help to determine increase and decrease price up to a certain limit.

Our model that predicts ads that takes into account factors like age, gender, marital status, current designation, work location, home location, travelling record, photo libraries and search terms, anniversary. Ads that are more specific to individuals are more profitable to publisher and satisfying to consumer.

**ACKNOWLEDGEMENT**

We owe a great many thanks to the following people who helped and supported us in the preparation of this project. We would like to thank our Director **Dr. C. MUTHAMIZHCEL VAN**, faculty of Engineering and Technology, SRM Institute of Science and Technology, Chennai.

We would like to express our deep sense of gratitude to **Dr. B. AMUTHA**, Professor and Head of Department of Computer Science and Engineering, SRM Institute of Science and Technology for giving us the opportunity to take up the project.

We would also like to thank my project guide **Dr. ANNAPOORANI K**,Professor, Department of Computer Science and Engineering for all the valuable guidance during the course of this project. We would like to thank all the teaching staff in the mini-project for immense support and valuable suggestions during the project reviews and course of this project.

We would like to take this opportunity to thank the SRM Management for their valuable support and express our heartiest gratitude to them.

Saransh Agarwal RA1511003010320

Mayuri Rani RA1511003010298

Palak Gupta RA1511003010292

**TABLE OF CONTENTS**

**ABSTRACT……………………………………………………………………….....iii**

**ACKNOWLEDGMET…………………………………………………...……...…..iv**

**LIST OF TABLES……………………………………………………...……...…….vi**

**LIST OF FIGURES…………………………………….……………..……...……..vii**

1. **INTRODUCTION**
   1. General trivia…………………………………………………………...…………..1
   2. Factors affecting todays advertisement…………………………………………….1
   3. Time based and Click based………………………………………………………..2
   4. Machine Learning in advertising…………………………………………………...2
      1. Data Mining…………………………………………………………………2
2. **EXISTING MODEL**
   1. General Trivia……………………………………………………………………...4
   2. What it is?.................................................................................................................4
      1. The algorithm……………………………………………………………….4
3. **DESIGN**
   1. Survey……………………………………………………………….…...…….…..7

3.1.1 Why it is used……………………………………….…………….....….…..7

* + 1. Implementation ………………………..……..…………………,……...…..8

3.1.3 Result of survey………………………………………………….……….....8

* 1. Data Mining………………………………………………………………..…….....9
     1. Why is it important……………………………………...……………….....11

3.2.2 Implementation ………………………………………………………...…..11

1. **IMPLEMENTATION**
   1. PHP and Html………………..…………………………………………………....13
   2. Signup page……………………………………………………………..…...……14
      1. Login Page…………………………………………………………….……..15
   3. Machine Learning…………………………………………………….…………..16
      1. Why is it important…………………………………………….…………..18
      2. Labelling…………………………………………………………….….….18
      3. Implemented work…………………………………………………..……..19
2. **SUMMARY………………………..........…………………………………………….21**
3. **APPENDICES…………………………………………………………………… ….24**
4. **REFERENCES………………………………………………………………………..25**

**LIST OF TABLES**

Table 3.1 Survey raw data from google form

Table 4.1 Signup Database

Table 4.2 Machine Learning features and label

**LIST OF FIGURES**

Figure 2.1 Old model for price manipulation

Figure 3.1 List of questions asked in the survey

Figure 3.2 List of attributes for the survey

Figure 3.3 Conversion from survey to usable data table

Figure 3.4 Conversion from Multiple attributes to single label

Figure 4.1 Website flow and structure of the system

Figure 4.2 Signup Attributes

Figure 4.3 Login Page

Figure 4.4 Machine Learning working process

Figure 4.5 Decision Tree Structure

**CHAPTER 1**

**INTRODUCTION**

* 1. **General Trivia**

Online advertising, also called online marketing or Internet advertising or web advertising, is a form of [marketing](https://en.wikipedia.org/wiki/Marketing) and advertising which uses the Internet to deliver promotional marketing messages to consumers. Web banners or banner ads typically are graphical ads displayed within a web page. Many banner ads are delivered by a central ad server. Banner ads can use rich media to incorporate video, audio, animations, buttons, forms, or other interactive elements using Java applets, HTML5, Adobe Flash, and other programs. Display advertising conveys its advertising message visually using text, logos, animations, videos, photographs, or other graphics. Display advertisers frequently target users with particular traits to increase the ads' effect.

**1.2 Factors affecting todays advertisement**

The factors that manipulates the prices of the product depends upon demand which in turn is basically dependent on a click rate and time together. But Other factors that defines todays advertisement include user’s preferences that is derived using machine learning and demand and market economy. But our study focusses on some new factors that is to let the Publisher decide the price at which he wants to advertise the product. Another factor is orientation based suggestions that determines what kind of ad a person will se. This is an example where males are shown more male oriented ads and females are presented with more of female oriented ads.

One more thing that is to be done is to stop advertisements that have already been bought by the consumer. This will in turn impact the perception of the ads suggestions to the consumer. By not showing ads for already bought items makes the consumer more satisfied that the price at which he bought the item is not more than what is advertised now. This is to minimize the regret of the buyer.

**1.3 Time Based and Click Based**

As previously discussed, letting the Publisher decide any change in price to occur or not gives the Publisher some control over the price at which his product is sold. Traditionally, the Publisher and Advertiser (ads space) doesn’t have any control over the pricing or even the ads that are going to be shown on their spaces respectively. This is done heuristically. This approach gives better results than traditional methods of pricing a product. This approach gives better results than traditional methods of pricing a product. In this method, the prices of product are regulated using time base and counter based. In time based the prices are increased as time passes and is also based on number of clicks. These both factors help to determine increase and decrease price up to a certain limit.

**1.4 Machine Learning in advertising**

Some other methods for Ads suggestions are having data of the consumer like his any anniversary date and showing him/her relevant Ads. These personalized ads make more sense and reduce any negative response from consumers. Facebook and Google, or even LinkedIn’s data can be helpful for letting Ad services to personalize ads for each consumer for various occasion and Moods. Like Facebook’s statuses can be used to suggest ads like some music lover may be suggested new pair of headphones or a person with more number of athletic photos can be shown more ads for sports equipment. Although Ads services are already smart but the limitation is that prediction based ads. One example of predictive ads is a person searching for engineering colleges on google can be shown ads for laptop or a smart phone because engineering students generally buy new laptops and smartphones at the start of their engineering. Laptop ads are not shown to the consumer until he/she searches for it on google or amazon. This is a huge limitation because this suggest that ads are merely based on some search keys but not on intelligence.

**1.4.1 Data Mining**

Improved data-mining’s predictive algorithms (Machine learning) can be used to create models that predicts ads that takes into account factors like age, gender, marital status, current designation, work location, home location, travelling record, photo libraries and search terms, anniversary. Ads that are more specific to individuals are more profitable to publisher and satisfying to consumer because these ads contain more probability for them to be sold and make profit for the publisher and also consumer specific ads based on all factors mentioned above that are relevant to the consumer eases the effort for the consumer to go search for them and makes it easy for the consumer to find them and buy them.

The ADs based on consumers interest is shown according to machine learning algorithm. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules. A survey has been taken to determine the interest of the person based on age, gender, marital status and designation. Age, Gender, marital status and designation are the levels of the features of the tree and the output data this is the ADs are the labels. This data is then fed in the machine learning to take out the desired output for the newer set of input.

**CHAPTER 2**

**EXISTING MODEL**

**2.1 General trivia**

The trend of selling ads based on time, rather than click-through rate, is gaining momentum

The strategy is meant to combat the wider problem of viewable ads. Approximately half of online ads sold are never viewed, causing massive waste in advertiser budgets. According to analytics company Moat, only 56 percent of ads on desktop are classified as in view; this drops to 45 percent on mobile. Rather than forcing advertisers to create bigger, flashier and more interruptive ads, publishers are charging for the ad only if it caught the reader’s attention.

The AD exchange industry is time based and counter based. The project deals with the functions of ad exchange industry. The ads are shown on the various websites according to users demand and preferences. The user buys the product by seeing the ad on the website.

**2.2 What it is?**

The user first sign ups the page then login in the website. The Ads are shown based on machine learning. A survey was taken. The labels obtained from the survey is fetched in the machine. The output is generated based on the decision tree. The output is then then shown as the Ads to the viewer. It shows the most relevant ads. The price of the Ads is set according to the publisher’s demand. This price is a variable. Its keeps on increasing and decreasing based on the algorithm set. The price increases up to a certain limit and decreases up to a certain limit.

**2.2.1 The algorithm**

According to the algorithm, the price varies based on number of clicks and the time the ad is viewed. The user views the Ad and then clicks on it in order to get the full glance of the Ads. The initial price is set the first time the user clicks on the ad. This is the starting price. Further, if the user again clicks on the ad the counter file counts the number of clicks.

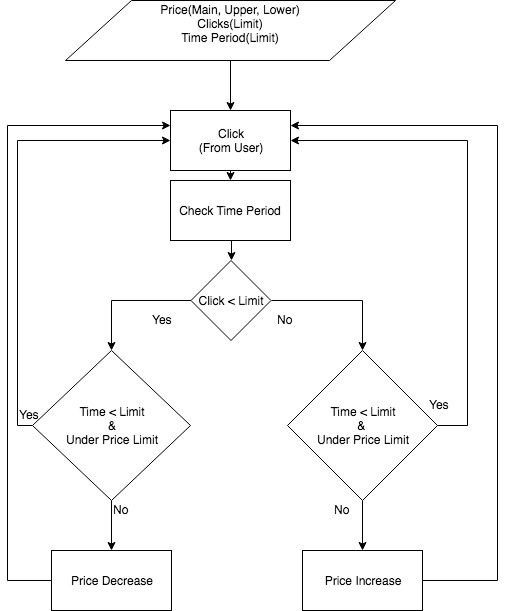


Figure 2.1 Old Model for price manipulation

The current time is stored as soon as the user login to the page. The time interval is calculated thereafter. In real time world, the time interval may last up to few days or weeks based on the publisher’s demand. But to implement in our project we took it as few seconds to show the implemented working model.

The user clicks on the ad and the counter value is incremented. The time interval is continuously checked. If the time interval exceeds a certain time period then the price is increased based on the number of clicks. If the number of clicks exceeds a certain an amount then the price gets increased. and on the other hand, if the clicks are less than the certain amount then then price gets decreased.

This factor is used to keep the user interest in the product. If the clicks are less than the price is decreased so that it pulls the users interest to the lower priced product easily. On the other side, if the clicks are more than the user is interested in the ad. This increases the price so as to satisfy the publishers demand. This helps the buyer and publisher both to stay in no regret state. And thus, increase the smart advertising.

**CHAPTER 3**

**DESIGN**

**3.1 Survey**

Surveys are a method of gathering information from individuals. Surveys have a variety of purposes, and can be conducted in many ways. This information is collected through use of standardized procedures so that every participant is asked the same questions in the same way. It involves asking people for information in some structured format. Depending on what is being analyzed, the participants being surveyed may be representing themselves, their employer, or some organization to which they belong.

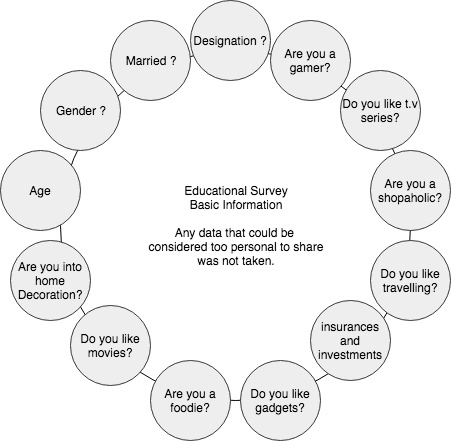


Figure 3.1 List of questions asked in survey.

**3.1.1 Why it is used?**

A survey can be used to investigate the characteristics, behaviors, or opinions of a group of people. These research tools can be used to ask questions about demographic information about characteristics such as sex, religion, ethnicity, and income. They can also collect information on experiences, opinions, and even hypothetical scenarios. For example, researchers might present people with a possible scenario and then ask them how they might respond in that situation.

**3.1.2 Implementation**

In order to achieve data for different sets of people, we also performed a survey on different sets of people. Through this information about users were gathered. Questions were based on likes and dislikes of a person such as preferences for gaming, shopping, gadget, home decoration, movies, food, travelling. It is done in order to create training set to perform data mining.

Survey was performed by creating a Google form where general information about a person where asked. Survey link was circulated among college students, Senior citizens, Working people and youngsters. Questions asked where simple yes and no answer questions related to persons likes and dislikes such preferences for gaming, shopping, gadget, home decoration, movies, food, travelling.

**3.1.3 Result of Survey**

We received a total number of one hundred and six (106) responses for the form. Out of these responses two (2) responses were from age group of eighteen (18), three (3) from age group of nineteen (19), twenty nine (29) from age group of twenty (20), thirty seven (37) from age group of twenty one (21), twelve (12) from age group of twenty two (22), seven (7) from age group of twenty three (23), one (1) from age group of twenty four (24), two (2) from age group of twenty five (25), one (1) from age group of twenty nine (29), two (2) from age group of thirty four (34), one (1) from age group of forty (40), one (1) from age group of forty four (44), two (2) from age group of forty five (45), two (2) from age group of forty six (46), one (1) from age group of forty eight (48), two (2) from age group of forty eight (48), one (1) from age group of fifty two (52).

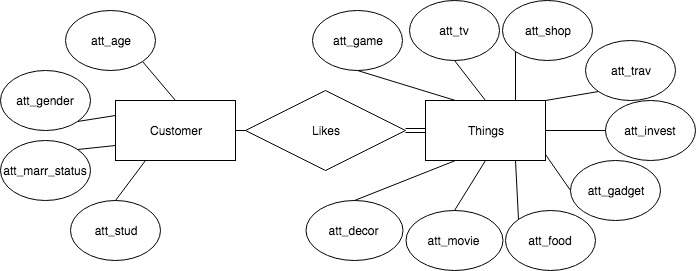
****

Figure 3.2 List of attributes for the survey.

With respect to gender, seventy-five (75) are males; thirty-one (31) are females. Ninety-four (94) of the responders are not married and twelve (12) are married. Ninety-three (93) were unemployed and thirteen (13) were employed. Sixty (60) of the responders liked playing games whereas rest forty-six (46) responders does not play games. Seventy-six (76) of the responder’s love watching TV series and other twenty-seven responders didn’t like watching TV series. Sixty-six (66) of the responder’s love shopping and rest forty responders do not like shopping. With respect to Travelling, ninety-six (96) of the responder’s love travelling and other ten (10) does not like travelling. In accordance to insurances and investments Fifty-one responders are interested in investments and rest fifty-five (55) are not interested in it. The result generated would be fed as the training set for data mining. Using this information predictions for different sets of people will be generated so that user centric ads could be displayed to the user thereby increasing the efficiency as well as productivity and total profit of the Seller.

**3.2 Data Mining**

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. In data mining, association rules are created by analyzing data for frequent if/then patterns, then using the support and confidence criteria to locate the most important relationships within the data. Support is how frequently the items appear in the database, while confidence is the number of times if/then statements are accurate. Data mining techniques are used in many research areas, including mathematics, cybernetics, genetics and marketing. While data mining techniques are a means to drive efficiencies, and predict customer behavior, if used correctly, a business can set itself apart from its competition through the use of predictive analysis.

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. It is an essential process where intelligent methods are applied to extract data patterns. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.



Figure 3.3 Conversion from survey to usable data tables

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

**3.2.1 Why is it important**

In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact businesses. Data mining is an important process to discover knowledge about your customer behaviour towards your business offerings. It explores the unknown credible patterns those are significant for business success.

With data mining, Business Organizations are able to make more accurate business decisions and incur more profits. From business, marketing advertising and introduction of new products or services, and everything in between. Data mining draws the results to:

• Improve customer loyalty

• Find hidden profitability

• Reduce Client Churn

**3.2.2 Implementation**

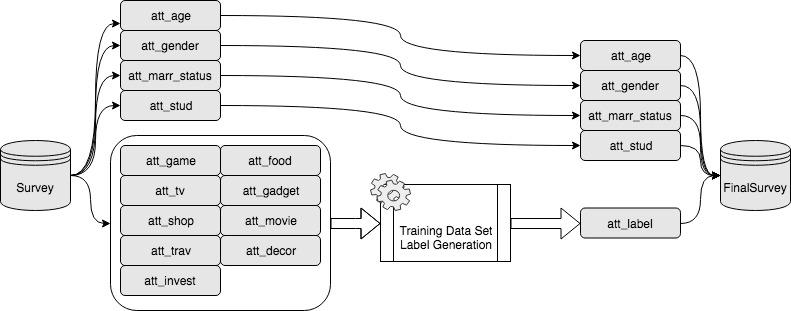
****

Figure 3.4 Conversion from multiple attributes to single label

The algorithm implemented for mining data is decision tree algorithm. Its goal is to create a model that predicts the value of a target variable based on several input variables. Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). Data is collected by performing a survey. Implementation of this survey is discussed in the previous chapter. With of data mining predictions for future sets of data are deduced. Features used to describe user in our project are age, gender, marital status and designation.

The obtained data are converted in 1 and 0 form in order to feed it in for machine learning.

****

Table 3.1 Survey raw data from Google form.

This data was later converted to integer values from “Yes” to 1 and from “No” to 0. The integer values are easy to manipulate in terms of implementation using decision tree in machine learning which is a completely dependent on integers as deciding factor to categories.

**CHAPTER 4**

**IMPLEMENTATION**

**4.1 PHP And HTML**

Our Model has implemented Machine Learning to produce user-centric ads. This is achieved by performing a survey. Features that are used to describe user are age, gender, marital status, designation. Machine learning helps in producing ads for user that produce effective results. Main Programming Languages used for implementing our program are PHP, HTML. SQLite3 is used to maintain the database.

The price of products are kept fluctuating such that the efficiency is increased.PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. PHP code may be embedded into HTML code, or it can be used in combination with various web template systems, web content management systems, and web frameworks.

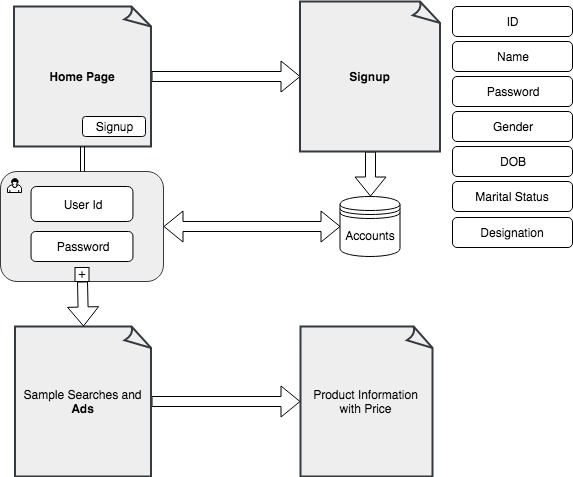


Figure 4.1 Website flow and structure of the system.

The entire infrastructure of our project is built upon PHP and HTML. HTML is used to develop signup and login page. CSS is also implemented to make the site more attractive. Our Model has implemented Machine Learning to produce user-centric ads. This is achieved by performing a survey. Features that are used to describe user are age, gender, marital status, designation. Machine learning helps in producing ads for user that produce effective results. SQLite3 is used to maintain the database. One of the reasons of using PHP for development is its fast performance. PHP also have wide Community due to which finding help is extremely. Other than this, PHP is easy to use and portable, that is, it could run on various platform thereby increasing the efficiency.

**4.2 Sign Up Page**

User Signup Page is means to add new members to a community or adding members to whom some unique information could be provided. Through Signup pages information about the members can also be gathered. Although, it should have kept in mind that this information should secured as they could be private to the user. The purpose for gathering this information is to get some insight about the members so as to know them to a certain extent in order to provide them information that would prove to be beneficial both ways.

User Signup is used to get the primary features of the users. Details are stored in database using SQLite3.SQLite3 is a relational database management system contained in a C programming library. SQLite3 is a popular choice as embedded database software for local/client storage in application software such as web browsers.

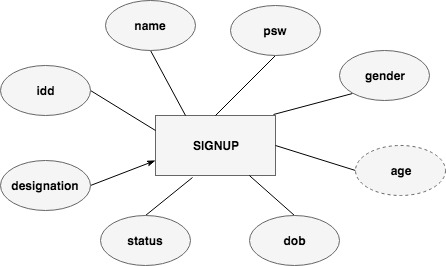


Figure 4.2 Signup Attributes

The signup page is developed using PHP and HTML. Information needed regarding user are asked to fed. The inputs are stored in a database. SQLite3 is the database that we have implemented in our project. Users are uniquely identified by allotting them a unique id. The information asked in the signup page serves as the feature set for our Machine learning algorithm. The feature set has values such as age, gender, marital status, designation. While requesting for information we have kept in mind to ask for information that aren’t private to user thereby not hindering their privacy.

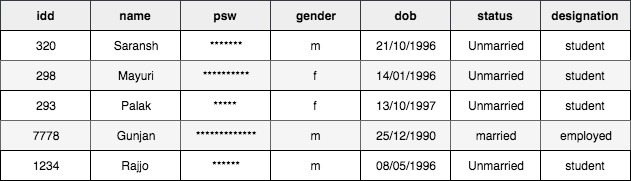


Table 4.1 Signup database

**4.2.1 Login Page**

A login Page is means to provide authentication that is, it’s a way to prove that the person is the one who he is telling us to be. Sign up page allows us to transfer confidential information around a circle without letting access to unauthorized persons.

Logging in is usually used to enter a specific page, which trespassers cannot see. Once the user is logged in, the login token may be used to track what actions the user has taken while connected to the site.

The login page in our project is developed using PHP and HTML. Members have to feed their primary credentials in order to access the page. The credential here asked are user’s unique id and password both set by the user. We have insured that no two members have the same ids thus preventing any conflicts in the identities of the users. Although, user could use any password he wants to. It’s totally up to user to decide a strong password as no such restrictions are provided from our side. If the provided credentials match a row in the database then access is granted to the user. Login Form is a way for providing authentication but there is no guarantee for confidentiality through this. Since our project concerns with showing user centric ads and to describe the working of our algorithm that’s why we have not looked into this matter.

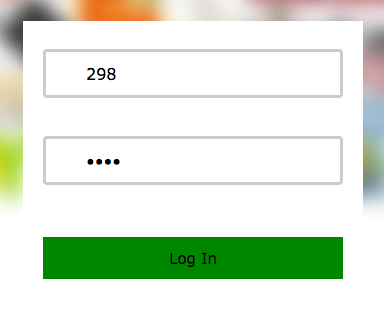


Figure 4.3 Login page

**4.3 Machine Learning**

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entitiesand then used to find meaningful anomalies.

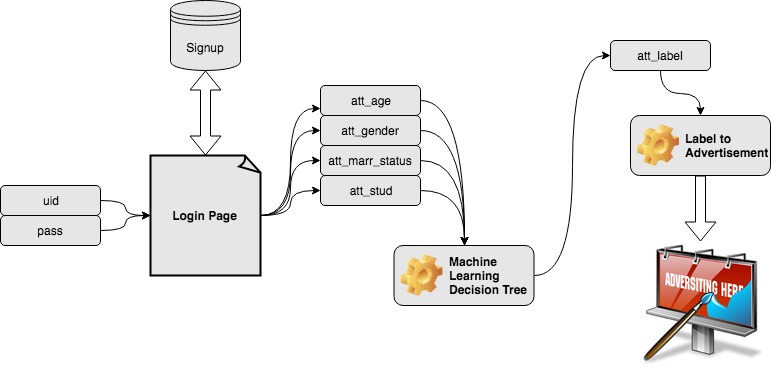


Figure 4.4 Machine Learning working process

A core objective of a learner is to generalize from its experience. Generalization in this context is the ability of a learning machine to perform accurately on new, unseen examples/tasks after having experienced a learning data set. The training examples come from some generally unknown probability distribution (considered representative of the space of occurrences) and the learner has to build a general model about this space that enables it to produce sufficiently accurate predictions in new cases.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory. Because training sets are finite and the future is uncertain, learning theory usually does not yield guarantees of the performance of algorithms. Instead, probabilistic bounds on the performance are quite common. The bias–variance decomposition is one way to quantify generalization error.

For the best performance in the context of generalization, the complexity of the hypothesis should match the complexity of the function underlying the data. If the hypothesis is less complex than the function, then the model has under fit the data. If the complexity of the model is increased in response, then the training error decreases. But if the hypothesis is too complex, then the model is subject to overfitting and generalization will be poorer.

**4.3.1 Why is it important?**

Machine learning has several very practical applications that drive the kind of real business results – such as time and money savings – that have the potential to dramatically impact the future of our organization. At Interactions in particular, we see tremendous impact occurring within the customer care industry, whereby machine learning is allowing people to get things done more quickly and efficiently. Through Virtual Assistant solutions, machine learning automates tasks that would otherwise need to be performed by a live agent – such as changing a password or checking an account balance. This frees up valuable agent time that can be used to focus on the kind of customer care that humans perform best: high touch, complicated decision-making that is not as easily handled by a machine. At Interactions, we further improve the process by eliminating the decision of whether a request should be sent to a human or a machine: unique Adaptive Understanding technology, the machine learns to be aware of its limitations, and bail out to humans when it has a low confidence in providing the correct solution.

Machine learning makes it possible to build systems which can handle tasks which cannot be programmed. Some tasks are simply impossible to code.

Machine learning not only allows computers to respond to tasks they have been taught, it allows computers to make leaps of intuition - to recognise tasks which are similar enough to previous experience to solve.

It’s very difficult to program intuition, because in a real-world situation it is not always possible to identify the key abstractions required to identify a situation - to separate information from noise. Machine learning removes the need to identify said abstractions - the machine figures them out for itself.

**4.3.2 Labelling**

Labelled data is a group of samples that have been tagged with one or more labels. Labelling typically takes a set of unlabeled data and augments each piece of that unlabeled data with meaningful tags that are informative.

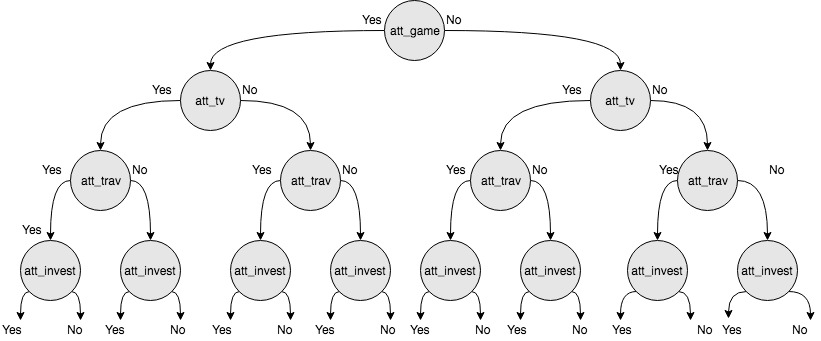


Figure 4.5 Decision Tree structure

Labels can be obtained by asking humans to make judgments about a given piece of unlabeled data, and are significantly more expensive to obtain than the raw unlabeled data. After obtaining a labelled dataset, machine learning models can be applied to the data so that new unlabeled data can be presented to the model and a likely label can be guessed or predicted for that piece of unlabeled data.

**4.3.3 Implemented work**

The main outcome of our project is through machine learning. Machine learning provides with the better result through datasets. The data is fed into the machine and labels or the outcome of the project is generated. The resulted data is used to calculate the Ads label required for our project.

A survey has been taken to analyze the interest of the people based on age, gender, marital status, designation. We have taken into consideration the interest and preferences of various sets of people in the field of gaming, movies, shopping, travelling, to series, foods, gadgets, home decoration and investments and funds. The survey is through circulated google forms.



Table 4.2 Machine Learning features and label

The output of the survey through database has been converted into two subgroups for machine learning namely features and labels. The features are broadly classified into age, gender, marital status and designation. The outcome of the interest of the various fields have been converted into labels and stored into database. This data is then fed into the machine as features and labels.

The machine is now trained for the newer sets of data. This will now help us to calculate the result for the people for their ad preferences when they signup the page. As soon as a user signup the page the data is sent to machine learning and output is calculated. When the user next login the page the ads are shown according to the calculated label. This way the user is more attracted to the ads as they are featured according to their preferences. This in turn helps the seller to also attract the user and increase their sells.

**CHAPTER 5**

**SUMMARY**

The main aim of our project is to give full satisfaction and bring involvement of the users for the Ads displayed in order to give both the user as well as the seller whose ads has been displayed in the website. This reduces the chance of regrets among the publishers and consumers. Machine learning and data mining has been used to estimate the result and to analyze as to which ads should be displayed in a particular user page. Different data sets and survey has been taken into consideration to arrive at the result set.

The older model was not efficient in displaying better ads. In the older model, only time and click based factors were taken into consideration to display the prices of the product. The ads were displayed according to gender. No more factors were taken into consideration. As a result, this led to very less productivity and biased ads. The users were not very satisfied with the ads displayed and hence were not interested in buying them. This led to the regret of the seller and insufficient use of the digital media to advertise the product for customer benefits. The incapability of the ads displayed without using machine learning and data mining led to tremendous loss in the ecommerce industry.

Later ads were randomized in order to give full fledge to all the field of industry. The users were shown ads according to the randomized algorithm and random ads were displayed irrespective of age, gender and interest. In the short term this proved to be good but when calculated for long term gain this idea was a complete fail. Peoples were not interested in all sorts of ads and some ads which should have not been user friendly are shown. For example, ads not based on age factor. Sometimes the randomized ads were not user interest centric and they were not interest in the particular ad and hence not giving any profit to the ecommerce industry.

Hence the shift to newer model was a necessity.

In the newer model, few concepts were introduced in addition to the older model. Machine learning and datamining are used in addition with the gender based ads and randomized ads.

The user first needs to sign up the page. The details of the user based on age, gender, marital status and designation is taken and sent to the machine learning algorithms in order to calculate the result for the ads to be displayed. The machine learning works on the survey which was taken.

The user also needs to login the page in python login page also. So that the details of the user are written in the file for the php script to read the users detail. The file then created contains the gender and the Machine Learning output label.

A survey has been taken to analyze the interest of the people based on age, gender, marital status, designation. We have taken into consideration the interest and preferences of various sets of people in the field of gaming, movies, shopping, travelling, tv series, foods, gadgets, home decoration and investments and funds. The survey is through circulated google forms. These details were used to form labels and features for the machine learning algorithm. The machine learning works on decision tree. The features are broadly classified into age, gender, marital status and designation. The outcome of the interest of the various fields have been converted into labels and stored into database.

Now when the user login the page the ads are shown according to the labels generated. In the ads page, there are two columns for ads display. On works on the machine learning algorithm and the other is gender centric. This helps us to categories the ads. The result is shown according to the survey. The result from the survey has been converted in binary form and stored in the database.

In the analysis of the ads to be displayed according to the labels, the labels of nine field of interest has been grouped into three sets. And then in these sets all the converted 0s and 1s are added. This as a result gives the labels in forms of one, two and three. Therefore, there are total three groups in a label. And further these labels are classified as – the label with number one shows the most preferred ad in that set according to the survey. The label with number two shows the ads based on gender. The label with number three shows the ads with randomized order in that particular set. Thus, the ads are displayed in order to fulfil all the requirements of the user. After the user is directed to the ads page, ads are shown. Now if the user clicks on the particular ad, he is directed to the details page. In this page, the product details are displayed. The price of the product is regulated on click and time basis.

The current time is stored as soon as the user login to the page. The time interval is calculated thereafter. In real time world, the time interval may last up to few days or weeks based on the publisher’s demand. But to implement in our project we took it as few seconds to show the implemented working model. The user clicks on the ad and the counter value is incremented. The time interval is continuously checked. If the time interval exceeds a certain time period then the price is increased based on the number of clicks. If the number of clicks exceeds a certain an amount then the price gets increased. and on the other hand, if the clicks are less than the certain amount then then price gets decreased.

Thus, this model helps in regulating the price and keeping in track the ads to be displayed. This binds the interest of the user as well as the seller.

**APPENDIX 1**

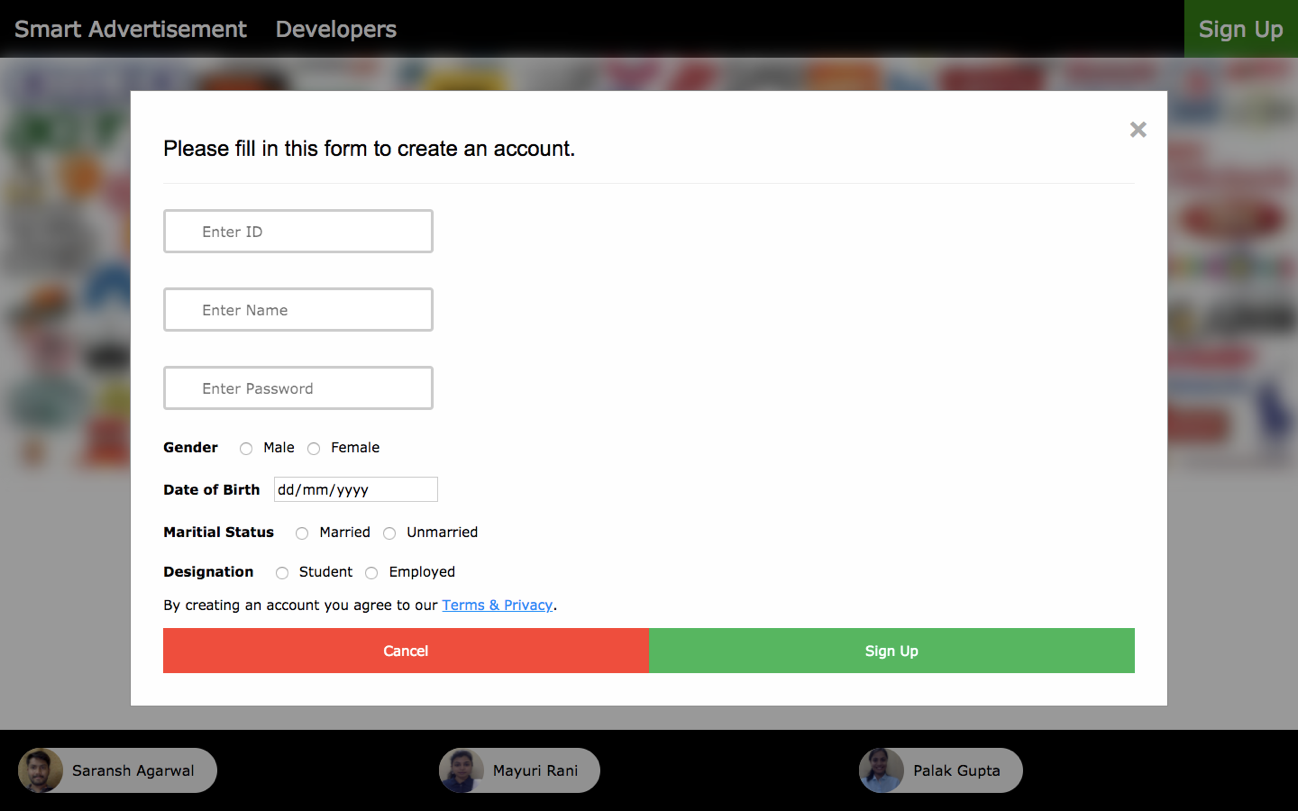
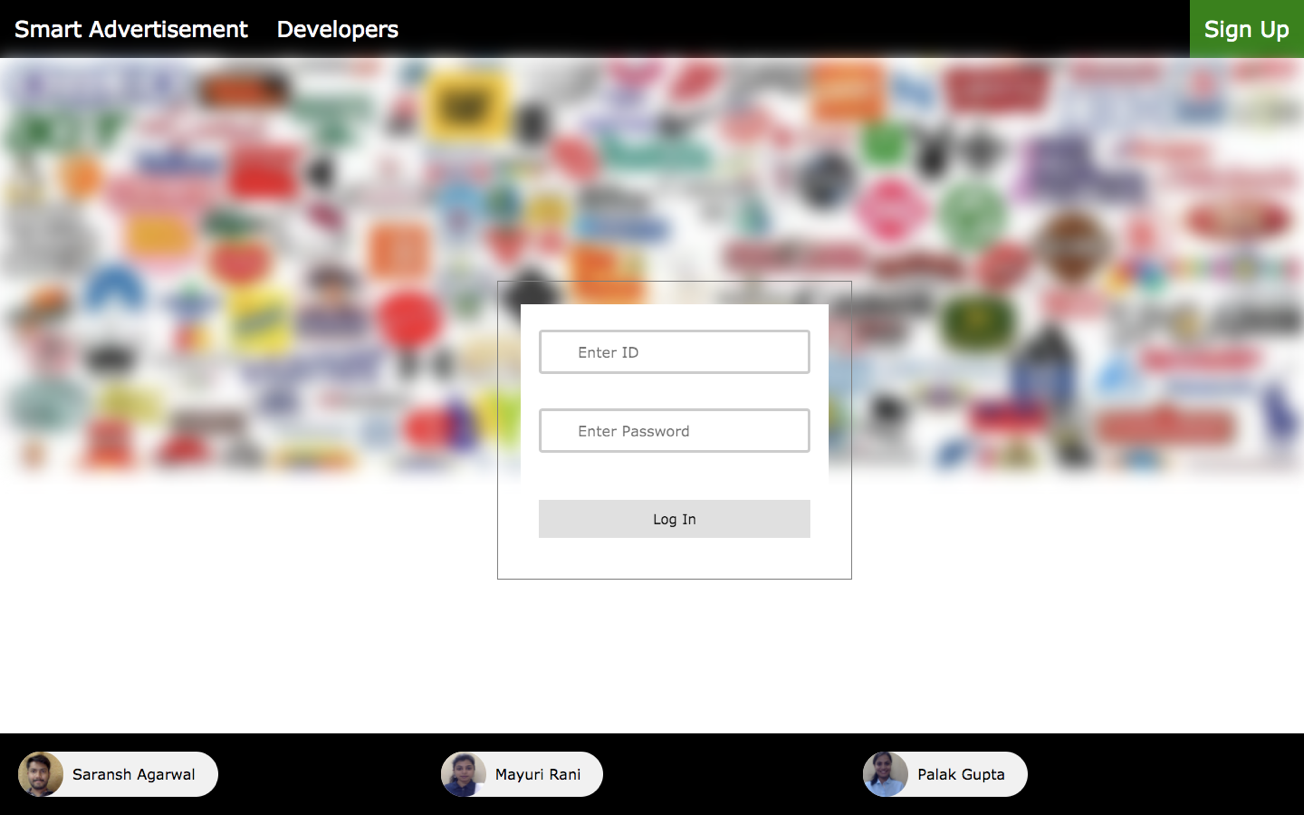


Figure A.1 User Signup Page

Figure A.2 Welcome Page

**REFERENCES**

1. Alessandro Acquisti, Hal R. Varian - “Conditioning Prices on Purchase History”– Vol. 24, No. 3, Summer 2005, pp. 367-281, ISSN 0732-2399

2. Rina Azoulay , Esther David –“Ad Exchange: Intention Driven Auction Mechanisms for Mediating Between Publishers and Advertisers “ - 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

3. Haibin Cheng, Erick Cantú-Paz -“Personalized Click Prediction in Sponsored Search” *WSDM’10,* February 4–6, 2010, New York City, New York, USA. Copyright 2010 ACM 978-1-60558-889-6/10/02 ...$10.00.

4. Matthew Richardson, Ewa Dominowska , Robert Ragno - “Predicting Clicks: Estimating the Click-Through Rate for New Ads” WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. - ACM 978-1-59593-654-7/07/0005.

5. Hoda Heidari, Mohammad Mahdian, Umar Syed, Sergei Vassilvitskii, Sadra Yazdanbod-“Pricing a Low-regret Seller” - *Proceedings of the* 33rd *International Conference on Machine Learning*, New York, NY, USA, 2016. JMLR: W&CP volume 48. Copyright 2016 by the author(s).

6. M. I. Jordan, and T. M. Mitchell-“Machine learning: Trends,- perspectives, and prospects” - 17 JULY 2015 • VOL 349 ISSUE 6245

7. Minping Zhou, Chaoyue Niu, Zhenzhe Zheng, Fan Wu, and Guihai Chen Shanghai Key Laboratory of Scalable Computing and Systems Shanghai Jiao Tong University, China - “An Efficient Privacy Preserving And Verifiable Online Auction Mechanism For Ad Exchanges”

8. Ziv Bar-Yossef Incentive, Kristen Hildrum, Felix Wu -Compatible Online Auctions for Digital Goods

9. Zhe Gao, Qigang Gao – “Ad-centric Model Discovery for predicting Ads' Click-through Rate-Procedia Computer Science” 19 (2013) 155 – 162