Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Real-Time, Targeted, Out-Of-Home Advertising with Dynamic Pricing

Catríona O'Driscoll
14310289

3rd May 2019

A Dissertation submitted in partial fulfilment
of the requirements for the degree of
Master in Computer Science

Declaration

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

Signed: ____________________ Date: ____________________

Abstract

Advertising is a constantly evolving method of marketing. The methods of advertising, for example; social media, newspaper, poster, television and internet, increase consistently, with some of these options almost at saturation point. This results in intrusive and ignorable adverts. Outdoor or Out-Of-Home (OOH) Advertising, with media such as billboards, posters and flyers, is the oldest method of mass communication. It still maintains some notable benefits over other, more modern media, such as the constant visibility, broad audience reach, and the inability of consumers to click away or turn off the advert. OOH advertising is a static method of advertising however, and cannot update reactively to differences in its audience. In more modern media, personalisation and targeting are frequently employed, where the audience is predicted and a relevant advert shown based on these audience changes. The more relevant an advert is to its audience, the more effective and thus valuable to the advertiser it is. With this added value, there is the opportunity to charge appropriately. Dynamic pricing has recently emerged where the relevance of the advert is reflected in the cost of the advertising time slot.

Work in the area of targeted adverts often focuses on mobile and online advertising, using data from applications, cookies, and social media. This results in quite effective personalised ads but issues such as feelings of intrusion and banner-blindness can occur. Similarly, text mining of geo-tagged social media posts has been used to make location-based audience predictions, but the use of social media brings up concerns such as exclusion. The idea of static pricing is also used in related works, or time-of-use tariffs, which lacks a correlation between the effectiveness of the advert and the price to display.

This dissertation explores the combination of the benefits of OOH advertising with more modern methods. A system is proposed for showing audience-personalised adverts on a digital billboard using real-time, location-based data, such as the temperature, weather and social class of the area, with a dynamic pricing model included. This collected data is used to train a neural network and predict the most relevant genre of advertising at any given time. Adverts with associated genre labels are added to the database, and, based on the predicted genre label's ranking in each advert's associated labels, the price of the advert will adjust for the time slot accordingly; increasing when the ranking is higher and decreasing with decreased relevance. The location-specific data is updated over set intervals and input into the neural network classifier. This classifier then predicts the next genre label, causing the pricing to change adaptively.

The neural network classifier is evaluated with a testing dataset which is made up of real-time, location-based data, as well as simulated data. The appropriate output labels are known, and compared to the ones predicted by the classifier. The precision, recall and F1-score are computed from this testing. The dynamic pricing model is tested through JUnit tests where adverts and labels are input with the expected results compared to the results created by the program. The results show that the decision-making model and pricing model work as they should. Further, the results demonstrate that they work in such a way that proves this proposed system results in personalised, location-based, real-time advertisements for the audience, as well as a dynamic and cost-effective pricing model for the advertisers.

Acknowledgements

I would like to thank my academic supervisor, Professor Siobhán Clarke for her guidance and support throughout the course of my dissertation.

I would also like to thank Andrei Palade who gave me valuable technical help and feedback.

I would like to acknowledge the Central Statistics Office for engaging with my dissertation and clarifying questions for me.

Finally I would like to thank my good friend Diarmuid Hickey and my family for their encouragement and willingness to proof-read my work.

Contents

List of Figures

List of Tables

1 Summary

This dissertation is set in the context of Advertising. It explores the benefits of modern advertising media and strategies, particularly in the form of targeting and personalisation, as well as the more traditional media, with Out-Of-Home advertising being the main are of focus. The implementation proposed in this dissertation takes in location-based data; such as weather, time, and season. This data is then used to train an Artificial Neural Network Classifier to determine the most relevant advert genre to show at any given time. A dynamic pricing model is applied to run concurrently with this prediction component to adjust the cost of showing an advert based on its relevance at any time; the more relevant the advert genre to the predicted relevant genre, the more expensive it is to run the advert. These price changes are used to reflect the effectiveness of an advert at any time based on the environment at any time around the billboard's location. The goal of this dissertation is to explore the possibility of combining the beneficial aspects of different advertising media into one system, with the addition of a pricing model that adapts to the changes in this system. This goal was achieved in the dissertation with a working implementation of a system that takes in a number of location-based data streams which are used to predict the most relevant advert genre. A successfully applied pricing model uses this predicted genre to alter the costs of displaying adverts stored in the system. Ultimately, this dissertation proves the possibility of a promising system with the desired factors integrated.

2 Introduction

2.1 Subject of Research

This dissertation explores the design and implementation of a system that uses location-based data to determine the appropriate genre of advert to show on a digital billboard. This is coupled with a dynamic pricing model to charge advertisers appropriately based on the relevance of the advert at a given time. This will result in more personalised and targeted adverts. It could also lead to increased consumer satisfaction (4), it could also lead to increased interaction and engagement in the audience and boost advertiser revenue and brand recognition in the area. Dynamic pricing can also be used to maximise revenue for companies, so will likely combine well with the targeting aspect of the implementation.

2.2 Motivation for Research Topic

Advertising is a rapidly evolving area, mandating a constant need for innovation in order to maximise the effectiveness of advertising campaigns. The benefits and more in-depth analysis of this topic of research can be seen in Chapter 2, but this dissertation aims to explore the combination of advertising media to combine strengths and minimise the associated weaknesses of each medium. Dynamic Out-Of-Home advertising, targeting and personalisation, data analysis, and pricing are the elements examined and combined in this implementation, with emphasis placed on integrating their strengths together.

2.3 Dissertation Overview

Chapter 2 provides the background of and context to the dissertation, as well as information on some of the technologies used in the implementation.

Chapter 3 contains the State of the Art, with discussion and analysis of previous approaches in this area.

Chapter 4 describes, in detail, the design and implementation of the system built for this dissertation.

Chapter 5 contains the evaluation process and results of the dissertation.

Chapter 6 contains the Conclusion, which reflects upon the dissertation and the implementation, as well as some discussion on possible future work on this topic.

3 Background

This chapter gives an insight into the areas researched before beginning this dissertation. Some analysis of relevant papers and their approaches is shown, and technologies used in the dissertation implementation are explained.

3.1 Areas of Research

The main areas of research in this dissertation were advertising; particularly Out-Of-Home advertising and online advertising, pricing models, and neural network classifiers.

3.1.1 Advertising

A maxim of advertising states that the average consumer needs to see something at least 10 times before taking action and making a purchase (5). This has led to companies putting emphasis on the quantity of adverts, and contemporary buyers being exposed to more than 3000 messages per day in some form (6). A lot of companies are exploring the benefits of online and social media advertising, with more than half of businesses planning to increase their investment in social media marketing (7) and internet advertising having generated more than $230 billion in revenue in 2017 (8). As an advertising medium that has seen so much money and development invested in the past number of years, we can see both its benefits and downfalls being uncovered. As personal and detailed data can be collected on the activity of online users, this medium allows for personalisation of adverts, and the targeting of consumers to ensure, as much as possible, that the adverts are shown to the right person at the right time. In a survey by K. O'Donnell and H. Cramer (4), it was shown that over half of the participants strongly agreed that they would like to see adverts personalised to them, and the 2016 Jivox Benchmark Report (9) that compared personalised to non-personalised adverts showed a huge increase in consumer engagement with personalisation, and an increase of 3 times the performance of the non-targeted adverts. These benefits significantly bolster the likelihood of success of an advertising campaign and so seem the most logical path to follow as a company. However, negative reactions and results to the influx of online advertising campaigns seen constantly by consumers have been noted. Complaints of intrusive and distracting adverts can damage a company's brand image, and even personalised adverts can seem intrusive if not carefully created. In a time of increased privacy concerns, this can make consumers feel unsettled about their data being collected and stored. Another issue that has arisen through online advertising, is the phenomenon of "Banner Blindness". Banner Blindness was a term coined to describe internet users seeming to intentionally avoid looking at, or not noticing banner adverts. In an eye-tracking study by M. Burke et al. (10), participants' initial eye movements suggested conscious efforts to avoid advertisements with the majority of fixations on banners taking place in the first eye movement, before the banner locations had been encoded by the participant.

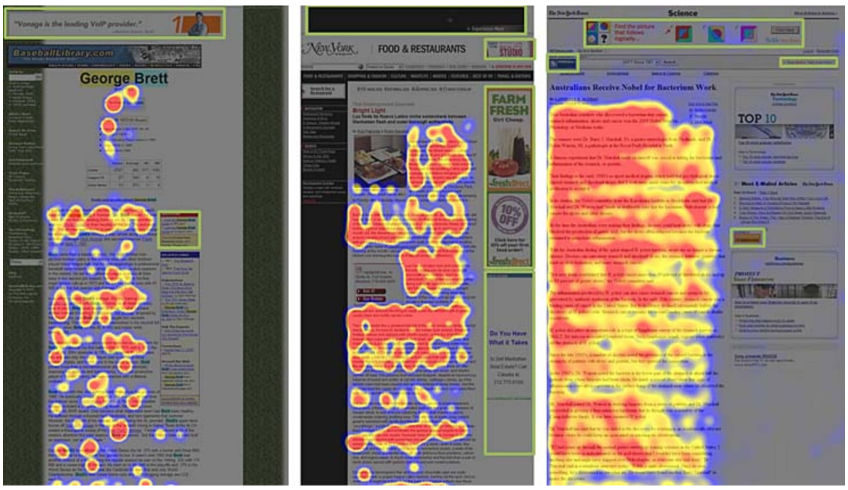


Figure 3.1: Heatmaps from user eyetracking studies by the Nielsen Norman Group (1). Red being the areas that users looked most, yellow areas having fewer views, blue being the least-viewed areas, and grey areas having attracted no fixations.

Outdoor, or Out-Of-Home (OOH), advertising is the oldest mass communication medium

(6) and includes forms of advertising such as billboards, street furniture, posters, flyers, and anything that would be visible to consumers while outside of their homes. Though now facing competition from modern advertising media such as search engines, social media and television, OOH advertising has grown steadily at a compounded annual growth rate of almost 15% (6). As such, it is still seen as a hugely beneficial method of advertising alongside the more modern media. In a study done by Nielsen and the Outdoor Advertising Association of America, over 50% of poster viewers are highly engaged, meaning they look at poster adverts either all or most of the time, and only 6% of the viewers responded that they "almost never" look at poster adverts (11). From this we can see that the most notable benefit of OOH advertising is the constant and noticeable visibility to large numbers of consumers, and of course, in comparison to online or digital advertising, it doesn't allow for consumers to click away or discard the advert. Within the realm of OOH advertising, there is development into Digital OOH (DOOH) options, where dynamic media is shown on OOH media. Of particular note are digital billboards which can show video and animation, as well as traditional posters, that can be changed and alternated throughout the day. These dynamic displays are becoming increasingly popular over static displays as more can be shown in the one place making advertising increasingly flexible, and allowing adverts to respond to changes in their audience (12), though typically restrained to time. DOOH incorporates elements of personalisation and targeting, but as a whole, the area of OOH advertising's most significant detraction is its static nature and inability to adapt to audience and environmental changes.

3.1.2 Pricing Models

Another area of research in this dissertation was pricing models, with particular attention paid to advertising. The basic pricing model is a static, flat tariff model where the price remains constant, independent of audience or location-based changes. There are also a number of dynamic pricing models that will be discussed in this work, and their applicability in the context of the dissertation. As a brief overview, dynamic pricing models are common in areas such as online advertising, smart grids (13) and digital content; areas where changes in supply, audiences and location are taken into consideration. Dynamic pricing models are considered effective tools to maximise revenues (14), and incentivise users to become more involved in payment plans and budgeting (13).

3.1.3 Artificial Neural Networks

The final piece of background information necessary before proceeding with the dissertation is an introduction to the concept of Artificial Neural Networks (ANN). ANNs are a framework in the area of Machine Learning that attempt to imitate the learning process of the brain. Biological neural networks have neurons that are connected by dendrites that receive inputs. Outputs based on these inputs are then sent through an axon to another neuron, and continue until a final output is reached and a decision made on initiating an action or not.

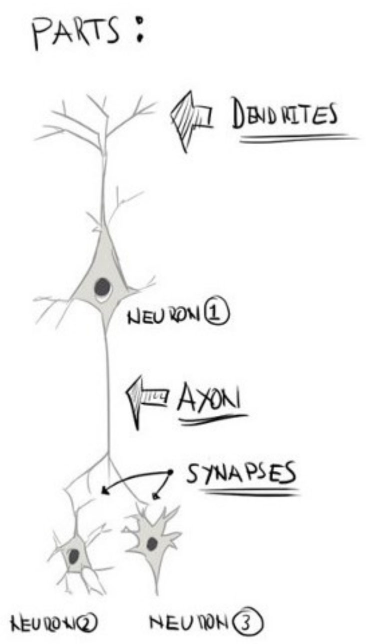


Figure 3.2: Depiction of biological neural network, with flow from top-to-bottom. (2)

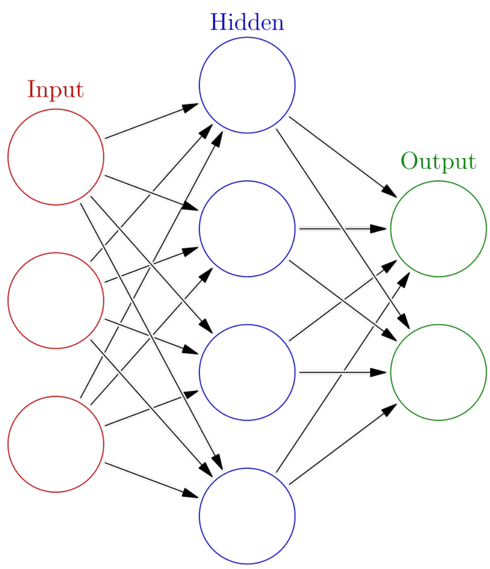


Figure 3.3: Artificial Neural Network, each circular node representing an artificial neuron. Flow from left-to-right (3)

In an ANN, layers of neurons are similarly used, taking multiple inputs and producing a single output to feed into the next layer of neurons. These systems 'learn' to perform tasks by being trained on examples, with no requirement or need for any task-specific rules. Weights are associated to each input and during training these weights are adjusted as they are compared to known outputs. This process is repeated, and weights adjusted, until the maximum number of iterations or an acceptable error rate has been met. ANNs are thus split into three layers: the input layer; which is given the input data to introduce to the deeper layers, the hidden layer; where neurons take in the weighted inputs, run a function appropriate for the needed output, and forward on the new output, and the output layer; which produces the new output from the model. The hidden layers are so-named as they do not directly interact with the data fed in or out of the network.

ANNs can adjust their parameters and learning process as new data is input, and are thus seen as highly flexible and adaptable in modelling. They have been used since the early 1980s in a large number of disciplines, such as engineering, medical diagnosis, data mining, meteorology and business (15), to interpret large datasets with multiple dimensions to come up with a high-level characterisation of the data (16). This classification aspect of ANNs is the application that will be used in the future in this dissertation. Classification being defined as "the grouping of things by shared features, characteristics and qualities" (17). ANNs are used in these classification problems where the relationship between the input and output appears vague, and large amounts of data are used to train the network in these scenarios so connections and weightings can be accurately computed by the ANN.

4 State of the Art

This chapter will present an overview and analysis of a selection of research papers and projects that have been published or proposed in the areas of advertising and pricing. They will be discussed within the context of the research topic, as well as the context of this dissertation.

4.1 Advertising

In the article "Branded cities: outdoor advertising, urban governance, and the outdoor media landscape" from 2012 (18), the author Kurt Iveson discusses trends within the advertising sector. Advances are noted in the scale, illumination, digital animation and sensory engagement of OOH adverts. Examples noted include;

- A Google patent of digital billboards that will be updated on local shop inventories;
- A London-based digital transit advertising system for buses that will enable screens to be updated as the bus moves through different locations; and
- An advert campaign in New York that would use audio technology and directional speakers above a billboard which can only be used by individuals as they pass through a spotlight location.

These examples demonstrate the significant effort being made in the industry to create impactful and relevant advertising campaigns while combining location and audience-based inputs in creative and diverse ways. They also highlight a stark contrast to a suggestion made in the book "Unleashing the Power of Digital Signage", written only 2 years previous (19), in which manually creating playlists of adverts was suggested as a method of showing relevant content to consumers. Though still used, in more recent papers and research, emphasis has been placed on creating more intricate combinations of data inputs and creative approaches. This "Branded Cities" article also discusses the morality and ethics behind partnerships between governments and advertising companies as a means of free urban infrastructure, such as bus stops and benches funded by advertisers. Of importance will be the monopolies these advertisers acquire as governments bring in regulations to decrease competing forms of advertising, such as telegraph pole posters and graffiti. These issues fall outside the parameters of this dissertation, but outline an important area of future study, addressing public-private partnerships in the future of OOH advertising.

4.1.1 Social Media

The concept of mapping urban areas, in this instance the cities of London and Barcelona, based on sensory data is explored in two papers by Schifanella, Aiello et al., "Smelly Maps: The Digital Life of Urban Smellscapes" (20), focusing smells of areas, and "Chatty Maps: Constructing sound maps of urban areas from social media data" (21), focusing on sounds. Both papers implemented a mapping of London and Barcelona using tagging information of georeferenced photos from social media sites such as Facebook, Twitter, Instagram and Flickr. Volunteers followed predefined routes and subsequently answered surveys about the sensory element, with these then being compared to the collected social media data to measure a level of accuracy. These papers do not approach these mappings with the explicit goal of including this data into advertising. However it offers a data analysis method which holds promise in a number of areas, including advertising. Smell and sound would be very useful within a combination of other inputs for advertising in an area, and these papers bring into question the idea of using this mapping approach to other types of inputs, which could also be included in advertising decisions.

Continuing with the social media approach, the article "Improved targeted outdoor advertising based on geotagged social media data" (12) proposes mining geotagged Tweets. This mining is done by using Latent Dirichlet Allocation, an unsupervised topic modelling method (topic models being a type of statistical model for discovering abstract topics that occur in a collection of data, frequently used for text-mining). In this article, Twitter data was text-mined from Underground rail stations in London, and themes found. The results found were mostly unsurprising, such as the fact that posts about "Sports" are most over-represented around stations near football stadiums and "Museums & Galleries" dominant in areas near large museums and art galleries. Being able to track these theme changes throughout the date would be hugely beneficial for targeting advertising. However, issues with data mining from social media exist. Twitter users, and social media users in general, are no representative of the wider population and an increase in Tweets or posts at a particular area does not necessarily match to a proportional increase in footfall in the area. Social media audiences, particularly Twitter, have been found to be over-represented by the younger adult population, interestingly with a noticeable leaning towards people of White-British ethnicity (22). This means that the diversity of the audience may not be correctly judged from this data and could constantly tend toward young audiences of a White British background, leaving the rest of the audience untargeted. Again, as an additional element of data to be analysed by advertisers, this is a useful avenue of research to develop but remains unreliable as the sole data stream.

4.1.2 Artificial Neural Networks

The next area of research was the inclusion of Artificial Neural Networks to aid in marketing and advertising. "Targeting direct marketing campaigns by neural networks" from 2011 (15) showed how the use of ANNs could better predict the response rate of recipients (would or would not buy) of direct mail marketing campaigns, thus improving the campaigns' effectiveness. The results from the ANN predictions were compared with the results of two conventional statistical models;

- Multiple Regression Analysis (MRA) - used to determine the relationship between one continuous dependent variable and two or more independent variables; and

- Linear Regression Analysis (LRA) - determines the relationship between a binary dependent variable and one or more independent variables.

This paper discusses the benefits and downfalls of the linear models as well as the ANN approach. In terms of input variables, the linear models can only take in a limited number and output a narrow range, whereas ANNs are very flexible with the number of possible inputs and outputs and can adapt and adjust the learning algorithm and parameters as needed. Linear models will give an estimated fixed-form function of relationships between the variables, unlike ANNs which are unable to define the relationships between their inputs and outputs. Ultimately this paper showed the ANN approach to be the most accurate and the effectiveness of a direct marketing campaign could indeed be improved with its use.

Further research is done on the topic of ANNs in advertising by Badhe Anup in 2015, in his article "Using Deep Learning Neural Networks To Find Best Performing Audience Segments". This paper looked at mobile advertising and the aggregation of a data profile on an advert. As the advert is clicked by a user data is collected, with this data being added to the associated data of the advert. The neural network is then fed this built up advert profile and determines appropriate audiences to show the advert to. This network then provides a binary output of 1 or 0 of whether the current user would click the advert or not. This article remarks on the complexities and pure size of data that is necessary to be collected to derive meaningful results and posits that ANNs are a definite way of optimising this process and revenue for the advertiser. This article also gave quite a detailed description of the process of data normalisation and the training of neural networks, which greatly aided in this dissertation.

4.1.3 Internet of Things

A developing area of technology is the Internet of Things, or Smart Devices, where devices can connect with each other, sending and receiving data. An example of this approach used in marketing was the Lexus 2015 advertising campaign. Smart billboards were set up at key locations in Australia, with cameras connected. These cameras would take a snapshot of the upcoming vehicles and communicate with a database to determine information such as the make, model and colour. From this data, targeted adverts would be shown with tag lines like "Hey black Merc driver, the heavens have opened. This is the new Lexus NX.", with information on the individual vehicles as well as some basic location-based data like weather. There is a debate ongoing on the regulations of billboards that can distract drivers, but as a method of creating eye-catching and personalised adverts, this was an impressive use of IoT in advertising.

In an article published at the 2016 International Conference on Innovation and Challenges in Cyber Security, titled "IoT based intelligent billboard using data mining" (23), an implementation of IoT with OOH advertising is proposed. In this paper, data analysis and mining is performed within a shop setting at the Point of Sale using RFID tags- a method of tracking using smart barcodes to identify objects. Using these tags it is possible to determine products left to sell and a digital billboard will prioritise offers on these products. This implementation offers the closest overall real-world model that could be found to that outlined in this dissertation. The weakness of this implementation, however, is that it is dependent on the use of RFID tags which are used in a limited number of areas, meaning it is shop-specific. It cannot be scaled-out to include other areas unless this tag becomes unnecessary, which would require an almost complete overhaul of the model. It also becomes aware of audience buying patterns after the purchases have been made and may not be necessarily indicative of future purchases. WIt nonetheless remains effective in its goal of emphasising bargains on remaining stock, but has a number of limitations once the aim is broadened.

"Advertising in the IoT era: Visions and challenges" from 2018 (24) discusses the issue of fragmentation within IoT and proposes a middleware to bring the hugely diverse range of Smart devices together into a common language and data collection layer. This proposal focuses on the idea of adverts within connected Smart cars. Profiles of the occupants of the car will have been generated through their respective Smart devices (eg. Mobile applications, online browsing on Smart phones) and these profiles will be sent to an intermediary platform between these end-users and advertisers. From these profiles, advertisers can determine which adverts to present through any number of the Smart devices within the car, potentially even the car itself - through the dashboard. This paper looks at the future of IoT as a method of approaching the area within the context of advertising. Fragmentation within IoT seems to be a core reason that this area is not currently being utilised so this aspect of research is interesting to see as a potential way of improving the implementation of this dissertation in the future.

Paper Name	Real-Time Adaptive	Scalable	Novel Approach	Cited by many
Branded Cities	No	Yes	No	50 - 60
Unleashing the Power of Digital Signage	No	No	No	50 - 60
Smelly/Chatty Maps	No	Yes	Yes	80 – 90/60-70
Improved Targeted Outdoor Advertising Based on Geotagged Social Media	Yes	Yes	No – social media mining is not a new avenue of research	~10
Targeting Direct Marketing Campaigns by Neural Networks	Yes	Yes	No – adds to work on ANNs	15 citation, 400 views
Using Deep Learning Neural Networks to Find Best Performing Audience Segments	Yes	Yes	No – adds to work on ANNs	30 - 35
IoT Based Intelligent Billboard Using Data Mining	Somewhat – adapts after purchase is made	No – requires RFID tags	Yes	2 citations, > 600 views
Advertising in the IoT Era: Visions and Challenges	Yes	Yes	Yes – A proposition of a future implementation	35 - 40

Table 4.1: Advertising State of the Art Summary

4.2 Pricing

Similar to the details of data analysis, pricing models and processes within advertising are not readily available to the general public. Official numbers and approaches are therefore largely unavailable. Though research papers offer some help in the topic of dynamic pricing as a whole, looking into current advertising companies and the ranges of values and small details they release tended to be of more use in the context of this dissertation.

4.2.1 Smart Grids

In the report "Dynamic response management of smart grids using dynamic pricing" from 2016 (13), the authors discuss a number of the common pricing options that are deployed. As mentioned in the Background Chapter, the most basic form of pricing is the static Flat-tariff model. In this model, the price remains constant regardless of any changes in demand or supply. This can lead to both overcharging and undercharging of the product (in this paper, the product being electricity supply), and losses to both the supplier and consumer. Dynamic tariff structures are presented in this paper as a method of flattening demand profiles and effectively planning electricity generation and distribution. There are a number of options put forward in this area;

- Block Rate Tariff: Consists of a number of tiers with different prices, comprised of consumers based on their consumption of electricity.
- Seasonal Tariff: As demand levels vary throughout seasons, higher rates are applied to the seasons with high demand, and lower rates to the seasons of low demand.
- Time-Of-Use (TOU) Tariff: The day is split into time segments. Higher rates are applied to time segments with high demand, and lowered for times of low demand.
- Real-Time Pricing (RTP): Prices change at regular intervals of one hour or less, to more accurately reflect the current situation.

The first three tariffs of this paper require historical data of consumers and their interactions with the smart grid. As such, it is difficult to apply from the beginning of a new implementation as this data does not yet exist. The RTP model is also the most relevant to the world of advertising as audiences are constantly changing and can be difficult to predict, in detail, in advance. This paper does however mention that this model is the one with the most uncertainty for consumers as it is constantly changing and not necessarily following a visible pattern of pricing, but increases the efficiency of the pricing scheme as it reflects the actual cost of supply. Within the context of Smart Grids in this paper, the authors believe that this would be the most difficult model to introduce to consumers, and the TOU or

Block Rate tariffs would have a smoother integration into the consumer market.

4.2.2 E-Commerce

The article "A Dynamic Pricing Model for E-commerce Based on Data Mining" (14) provides an insight into pricing within the area of e-commerce (the buying and selling of goods or services using the internet, and the process of the transaction of money and data). Similar to the previous paper, there are pricing models based on time, and dynamic adjustments based on a real-time assessment. The time-based model focuses on evaluating the different prices that the consumer can afford at different times. The dynamic marketing strategy then assesses the availability and inventory level of products to adjust the pricing, promotions and products being shown to consumers. This paper discusses the use of statistical models to process, mine, and analyse data. Within the analysis layer of the model, the authors mention the possibility of using association rules, classification, clustering, and sequential pattern analysis for decision-making. This approach requires data on the customers, and quite extensive research if the analysis and predictive element of this model is to operate with a high accuracy. It therefore offers an efficient approach for online media where this data is easily collected, but would be difficult, if not impossible to implement as a highly effective model in the area of Out-Of-Home advertising.

4.2.3 Advertising Pricing

A survey by The Digital Signage Insider (25) explored pricing of dynamic OOH advertising media by surveying suppliers in America. The concept of Cost-Per-Mille (CPM) - "M" being the Roman numeral for 1000 - is presented where the number of impressions or views of an advert is calculated to aid in pricing. CPM is calculated by dividing the total number of impressions by 1000. It is a method of determining the reach of an advertising campaign. From this survey it was seen that the majority (72.3%) of the respondents set their prices using this measurement as a major determining factor. In comparison, less than half (48.4%) indicated that specific viewer demographic characteristics were an integral factor, though it should be noted that determining these characteristics in verifiable and quantitative terms is generally accepted to be a very difficult task in the basic model of dynamic OOH advertising media (6). In terms of actual numbers in pricing, a range of $1-$10 was presented per thousand impressions, using the CPM method. This concept of CPM pricing is prevalent in online advertising as well where the number of impressions is quite accessible to measure with site visits and mouse-tracking software. By contrast, in the area of OOH advertising it is a measurement that is difficult to measure accurately. The common method of measuring footfall past the advert is somewhat rudimentary, as people can walk by without a glance at the advert, but still add to the advertiser's cost. Developments have been made in the area of cameras being attached to billboards and tracking crowd eye movements, as in the aforementioned Digital Signage Promotion Project in Tokyo, where data was collected on the number of people who turned their faces towards the billboard. This may be seen as somewhat intrusive on an audience ignorant of the data collection and is no longer seen as an accurate measurement in its current form in OOH media.

Research was also conducted on currently operating companies that maintain billboards and the pricing choices they have. Fliphound and Blip Billboards, both offering sites throughout the US, were the main sites of study. Fliphound offers an auction model on their billboards where pricing will be affected by the billboard location as well as the time of day that the advert is placed, due to traffic levels at a given time. The advertiser can also choose the frequency of their advert being shown from every 4 minutes, every 2 minutes, and every 1 minute, with the price increasing with frequency. Blip Billboards offers packages based on the advertisers daily budget and choice of location. Unfortunately, neither sites offer more detail in their pricing models and, though feedback is given to customers on estimated numbers of views of their advert, the exact factors and details in how these prices are determined are left inaccessible to the public. From these pricings some calculations, and assumptions on factor weightings, can be done to develop a pricing plan in this dissertation.

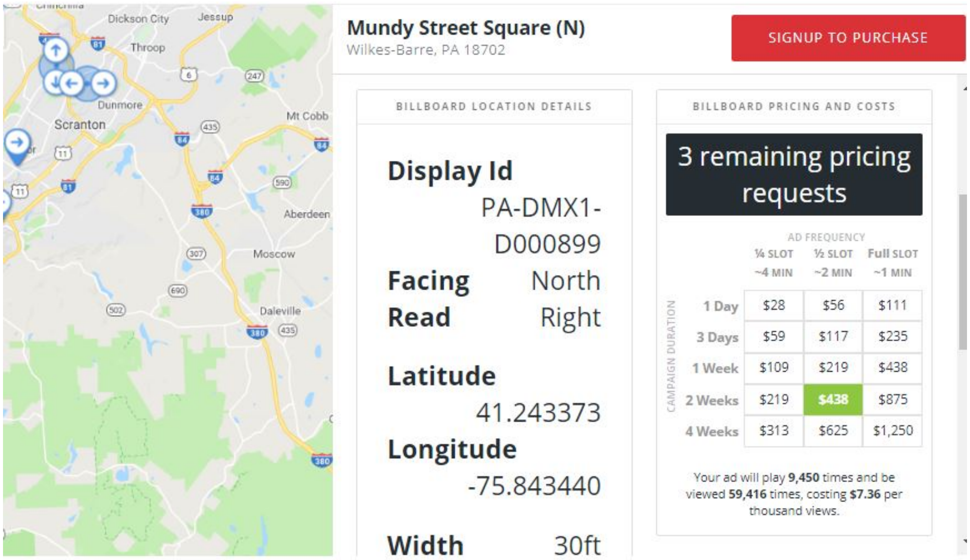


Figure 4.1: Fliphound Auction Interface

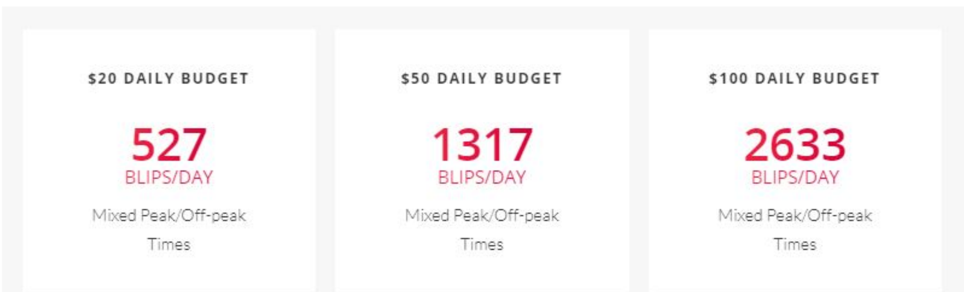


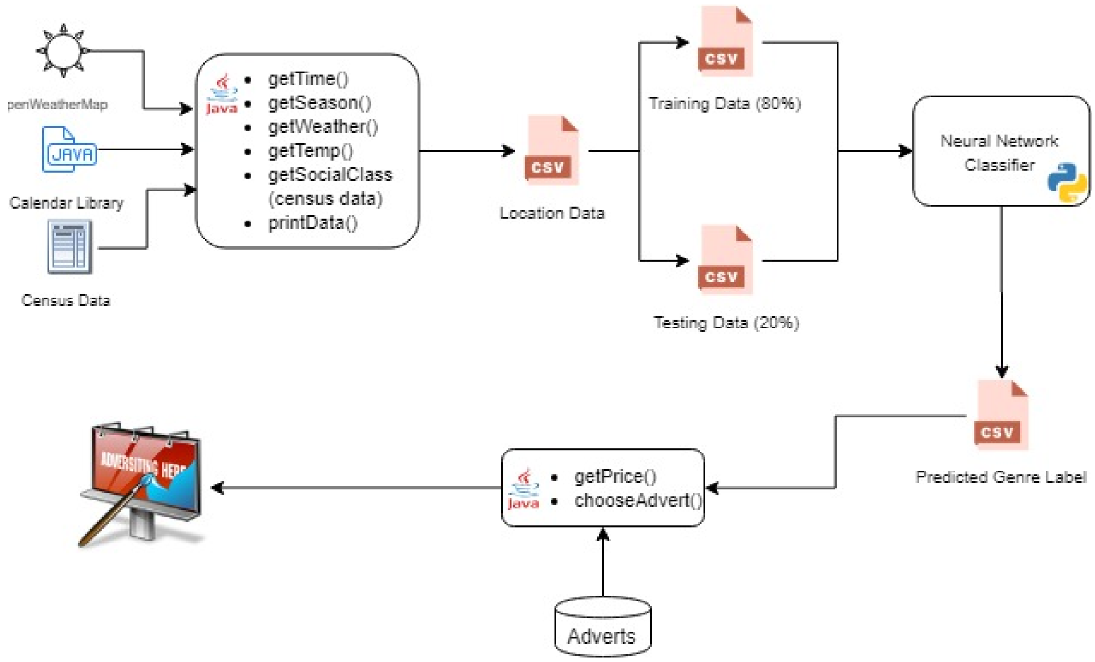
Figure 4.2: Blip Billboards Colorado Billboard Cost.

Paper Name	Real-Time	Measurements Needed Can Be Collected Accurately	Historical Data Needed	Can Be Used in OOH
Dynamic Response Management of Smart Grids Using Dynamic Pricing: Flat-Rate Tariff	No	Yes	No	Yes
Dynamic Response Management of Smart Grids Using Dynamic Pricing: Block Rate Tariff	No	Yes	Yes	Yes
Dynamic Response Management of Smart Grids Using Dynamic Pricing: Seasonal Tariff	No	Yes	Yes	Yes
Dynamic Response Management of Smart Grids Using Dynamic Pricing: Time-Of-Use Tariff	No	Yes	Yes	Yes
Dynamic Response Management of Smart Grids Using Dynamic Pricing: Real-Time Pricing	Yes	Yes	No	Yes
A Dynamic Pricing Model for E-commerce Based on Data Mining	Yes	Yes - Would be difficult	Yes	No
The Digital Signage Insider Survey on CPM	Yes	Somewhat	No	Yes

Figure 4.3: Advertising State of the Art Summary

5 Implementation

In this section of the dissertation, the design and implementation will be detailed and discussed, with explanations for the decisions made throughout the process. The image below lays out the final architectural design of the dissertation implementation on a high level.



The implementation can be broken down into three main components:

- Data Collection.
- Creation & Training of Neural Network Classifier.
- Development & Implementation of Dynamic Pricing Model.
- Scalability.

Each of these will be addressed in detail in this chapter.

5.1 Data Collection

Research was done on inputs that can affect the relevance and effectiveness of an advert at any time in a given location, and the ones chosen to be focused on were:

(1) Time

(2) Season

(3) Weather

(4) Temperature

(5) Social Class

(6) Nearby Events

(7) Nearby Places of Interest

A Java Program was created to collect and process all this location-based data, through a combination of methods which will be detailed below.

5.1.1 Time and Season

The priority of an effective advertising campaign is to reach the right person, with the right message, at the right time (26). The time of day has been shown to influence buying habits and, when accurately reflected in the advert, results in greater engagement, with research showing that 9am is the best time for luxury brand engagement online (26). The time of the year, or season, also has been shown to have an impact, with, for example, the first day of winter seeing people switch from salad to soup (27).

The time and season wer calculated every five minutes in this implementation, and were found using the Java Calendar Library. Using this library, the current time could be found in a 24-hour format. The day was then split into time-based slots: Similarly, from the Calendar

Table 5.1: Time Classification

12am-6am	Dawn
6am-9am	Morning
9am-12pm	Late Morning
12pm-2pm	Lunch
2pm-5pm	Afternoon
5pm-9pm	Evening
9pm-12am	Night

Library the month could be found using a Calendar instance and the *"get((Calendar.MONTH)+1)"* function, and the correct season (ie. Spring, Summer, Autumn, Winter) computed based on which month is returned.

5.1.2 Weather and Temperature

Weather and temperature values were collected using API calls to the OpenWeatherMap API. OpenWeatherMap is a digital weather information provider, established in 2014, and provides the current weather for any geolocation, as well as weather forecasts for up to two weeks in the future, and historical weather data for up to six years in the past. API calls are free for developers, though paid plans are available for more detailed requirements. This dissertation used the API to acquire the current weather of a given geolocation, in this case "Dublin, IE". This API call returns a Json file with a lot of information on the current weather at the time of call, shown in Figure 3.1.

This Json then has to be parsed to retrieve the relevant values from *weather:main* to get the weather description, and then *main:temp* to get the current temperature (as highlighted in Figure 3.1). The possible weather types are supplied by OpenWeatherMap at "https://openweathermap.org/weather-conditions", and are: Thunderstorm, Drizzle, Rain, Snow, Mist, Clear, Clouds. These can then be further broken down in the *weather:description* if required, to give details like "broken clouds", as seen in the sample response, or "light intensity drizzle", "heavy snow". In the context of this dissertation and advertising, the broader *weather:main* description level is sufficient.

```
{"coord":{"lon":-6.26,"lat":53.35},
"weather":[{"id":803,"main":"Clouds","description":"broken clouds","icon":"04n"}],
"base":"stations",
"main":{"temp":-2.89,"pressure":993,"humidity":85,"temp_min":-4,"temp_max":-2},
"visibility":10000, "wind":{"speed":1,"deg":190}, "clouds":{"all":75},
"dt":1548894600,
"sys":{"type":1,"id":1565,"message":0.0031,"country":"IE","sunrise":1548922306,
"sunset":1548954347},
"id":2964574,"name":"Dublin","cod":200}
```


Figure 5.1: Json Output from an OpenWeatherMap API call.

For ease of use, the temperature values, from *main:temp* were also categorised. These categorisations were created using various opinion pieces online (28), but as temperature feel per person is quite subjective, there have been some assumptions and broad, educated guesses made in this aspect.

Table 5.2: Temperature Classification

Extremely Freezing	(-30 <= Temp <= -10)
Freezing	(-10 < Temp <= 0)
Very Cold	(0 < Temp <= 10)
Cold	(10 < Temp <= 15)
Cool	(15 < Temp <= 20)
Warm	(20 < Temp <= 25)
Very Warm	(25 < Temp <= 30)
Hot	(30 < Temp <= 37)
Very Hot	(37 < Temp <= 50)

5.1.3 Social Class

Another aspect of an area that influences the type of adverts shown, is the social class, or socioeconomic status. Social class is the concept that the entire population is classified into specific socioeconomic groups. These classes being, as outlined by the Central Statistics Office (CSO) definition (29):

- Professional Workers;
- Managerial and Technical;

- Non-Manual;
- Skilled Manual;
- Semi-Skilled;
- Unskilled; and
- All other gainfully occupied and unknown.

These classes aim to include people with as similar as possible levels of occupational skill, with "All other gainfully occupied and unknown" used where no precise allocation is possible. In the area of advertising, differences in marketing campaigns can be seen in locations of different social class. For example, advertisements for fast food restaurants are more common in the most-disadvantaged areas, whereas advertisements for diet soft drinks, tea, and coffee are more common in the most-advantaged areas (30).

Social class can be computed using the census data of the location, provided by the CSO in Ireland. It is difficult to get averages of social classes, other than the high-earning and low-earning, using CSO data. To find a range and allow comparisons of different locations, the upper and lower bound of high and low-earners was set. For high-earners, the highest percentage in the country is in the Dún Laoghaire-Rathdown area at 55.9%, now set as the upper bound of high-earners, and the lowest bound being set to 27.7%, as found in Longford. Similarly, for low-earners the upper bound was set to the highest percentage in the country of 5.1%, from Monaghan, and the lower bound set to 1.3% from Dún Laoghaire-Rathdown. The census data of the location is then filtered and the number of Professional Workers and Managerial and Technical workers are added together as the "High-Earners" of the area, whereas the Unskilled value is set as the "Low-Earner" value of the area. The proportional ranking of the high-earners and low-earners is found by comparing with the range of lower and upper bounds for each status.

- If the high-earner ranking is over 75%, the area is classified as "Very High Earning"
- If between 50% and 75%, then "High Earning"

- If the low-earner ranking is between 50% and 75%, the area is "Low Earning"; and
- If the low-earner ranking is over 75% then the classification is "Very Low Earning".

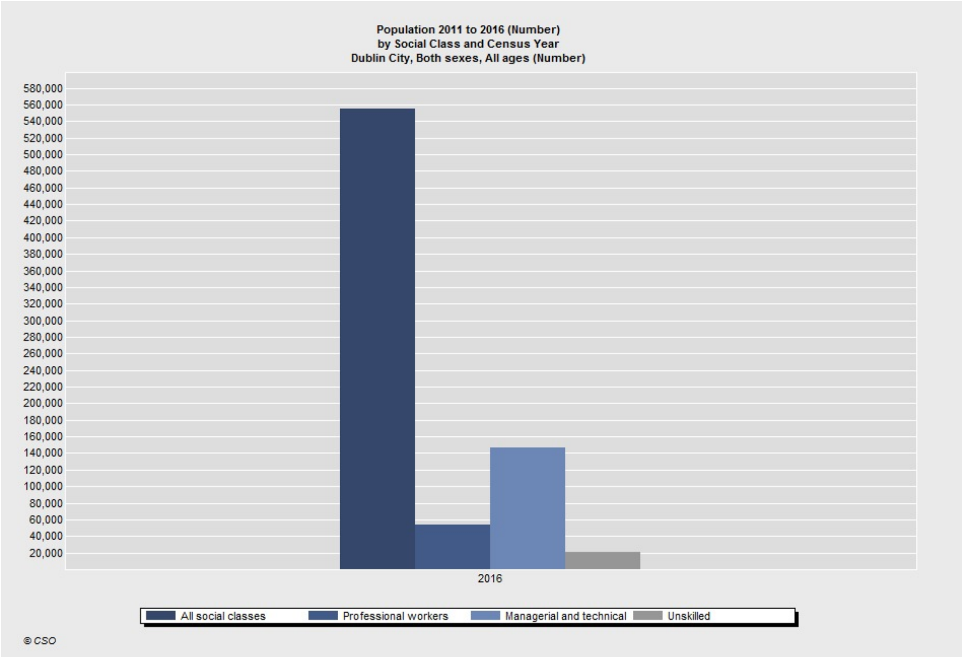


Figure 5.2: Dublin City Area Census with filtered information, 2016

An issue that arises with the census data is that it is only refreshed every four years, so may not offer up-to-date data streams being used. However this somewhat static nature means that once the location's social class is found, it need not be found again for a number of years. In this implementation, focus was kept on areas in Dublin so only census data CSV files from these areas were stored. However, the formatting remains constant throughout these files from the CSO, so to scale to include the rest of the country, it would just be necessary to include the remaining CSV files to the program.

5.1.4 Nearby Events and Places of Interest

An initial plan for this dissertation was to acquire information about upcoming or current events in the area using the Ticketmaster Developer API. The goal would be to extrapolate a genre of event and connect to a demographic class such as age, gender, or even taste in music. Though API calls could be made of events currently being shown on Ticketmaster in Dublin, further filtering of the API call or even the Json result proved incredibly difficult and time-consuming, and was put aside to continue with other possible avenues of research. Similarly with the Google Places API which, given GPS coordinates, would return nearby places of interest, the filtering and specification was very difficult and many helpful and relevant calls had just recently been deprecated and were no longer available. Unfortunately time constraints prevented the implementation of these data collection streams within this dissertation, but the above successful data streams gave enough data to continue with location-based targeting.

5.1.5 Design

With the data collection aspect completed, a connection needed to then be made to the adverts that are stored to show on a digital billboard. An Advert class was created which, in the initial stages, would be created with an associated season, weather, temperature, or area status. Based on the date then read in, an advert per data stream would be shown.

This initial decision-making model as shown in Figure 4.3 was the first step of testing the data incoming and ensuring it could be used to determine adverts to show. It was, however, much too basic, purely reliant on if, else-if statements, and unable to adapt to increased data streams or merge inputs. Though useful as a testing step, this decision-making approach was ultimately discarded and the implementation of a Neural Network decision-making model was decided upon. This decision followed further research on the topic of decision-making models, particularly in the area of advertising with multiple inputs and potentially outputs. This Machine Learning approach would replace the above, hard-coded decision-making model, and incorporate an intelligent and adaptable model to decide which advert to show at any given time.

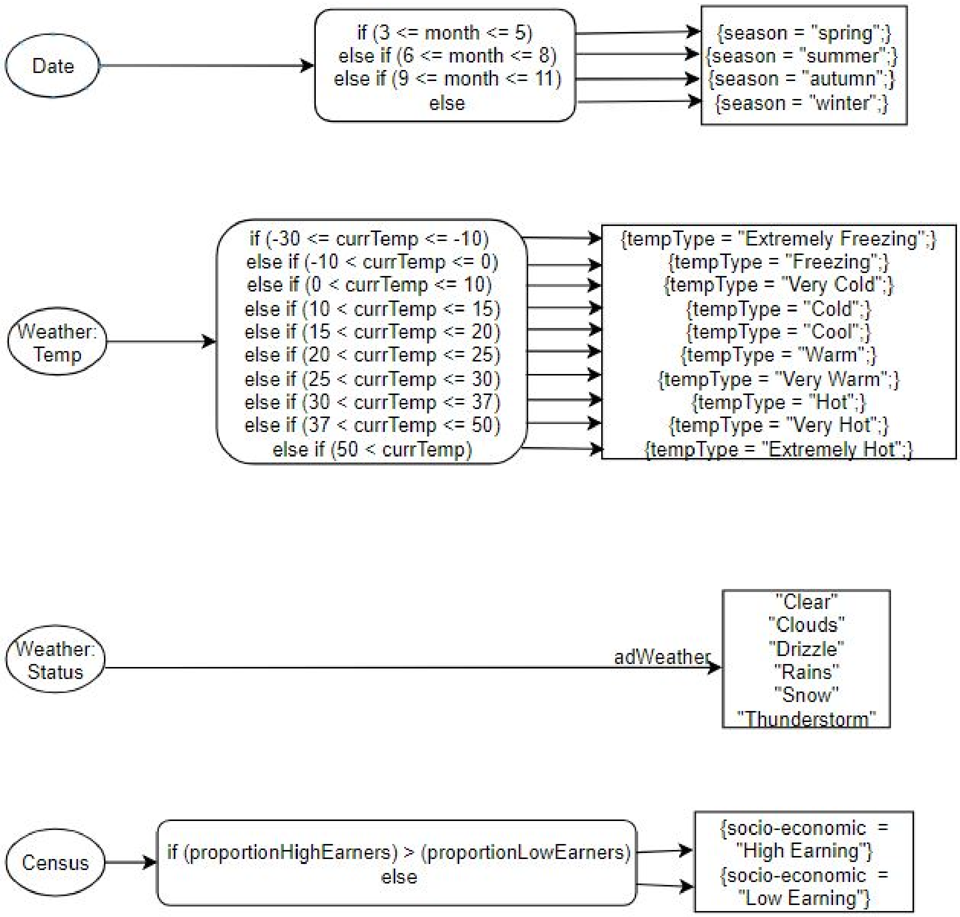


Figure 5.3: Initial Decision Making Model

5.2 Creation & Training of Neural Network Classifier.

Through thorough research of the area of Artificial Neural Networks (ANNs), an implementation of a neural network classifier became the logical application of ANNs to choose. ANNs in their basic form take in inputs, apply weightings, and then give a binary output, 1 or 0, to do the action or not. While this could be applied in the dissertation's context for each advert, with the output being to show the given advert or not, there would remain issues with multiple adverts being seen as relevant at a given time, and with just a binary association of relevance, there is no obvious way of ranking the "should-be-shown" adverts. An ANN classifier, in comparison, can take in multiple inputs and has a range of outputs to choose from. In the context of this dissertation, this application is more sensible as the outputs can be set as a collection of possible advert genre labels, which can then be compared with the adverts stored to rank relevance. For this dissertation the advert genre labels chosen, based on their frequent mentions during the research phase, were:

(1) Family

(2) Fast Food

(3) Luxury Brands

(4) Holidays

(5) Clothes

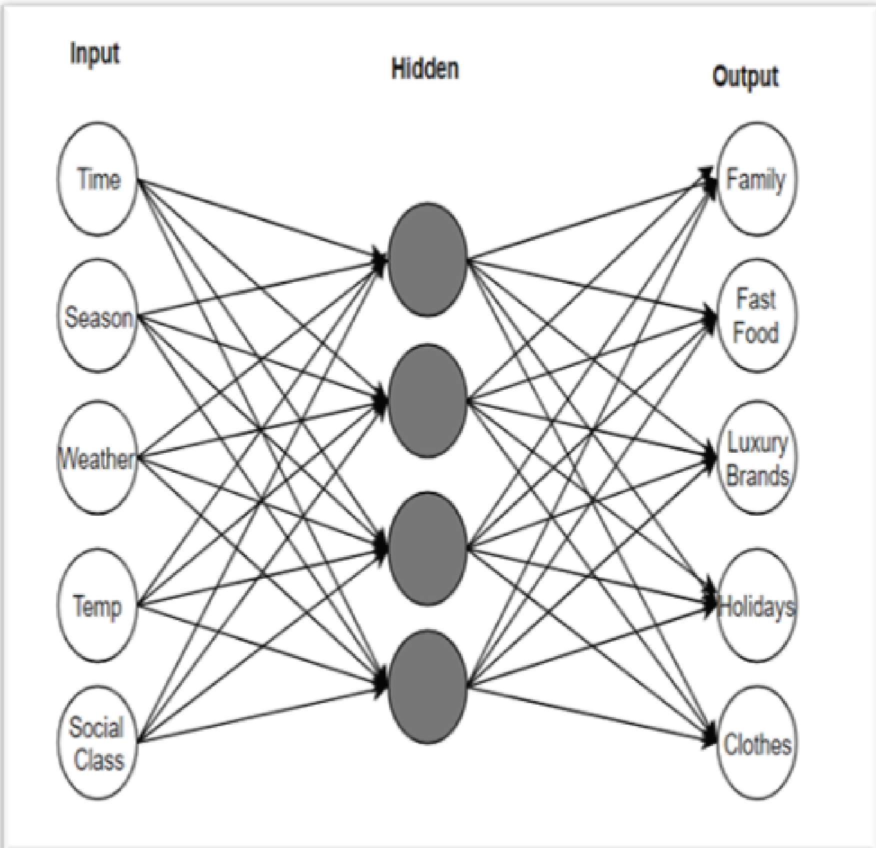


Figure 5.4: Visualisation of Neural Network Classifier model with inputs and outputs

5.2.1 Training & Testing Data

ANNs are self-learning and are trained to perform their function, in this case classification from input data, through repetition and sample, or training, data. To create this data, the data collection aspect of this dissertation ran for a several hours, collecting data every 5 minutes, and printing the values collected into a CSV file. Each line of input is given an output based on the values and so each genre label was analysed to see how the inputs affected them, with some assumptions made as advertising companies do not generally have their data analysis accessible for the general public. As the real-time data collected in the first step of this implementation was collected over a short range of time and does not cover all the ranges of the input values (for example; as the data was collected over a number of days in Winter, the temperature remained at a steady low value), simulated data was also included in the location data file to allow for a broader range of training and knowledge for the classifier.

Genre	Time	Season	Weather	Temperature	Social Class
Family	Morning - Afternoon	all	all	all	all
Fast Food	Late Morning - Night	all	Clear	Cool – Very Hot	Low – Very Low Earning
Luxury Brands	all	Spring & Summer	Clouds & Clear	Cool - Hot	High – Very High Earning
Holidays	Morning - Evening	Spring & Summer	all	all	all
Clothes	Morning - Afternoon	all	all (except Mist & Fog)	all	all

Figure 5.5: Genre labels with associated data values

Before this location data CSV file can be given to the classifier, the data must first be preprocessed, which can be done through Python. The file is first split into training and testing sets using SciKit Learn's *train_test_split* function. These sets must then be normalised as ANNs can be quite sensitive to non-scaled data. The same scaler must be applied to both the training and testing data, and again we use a SciKit Learn library, for the built-in StandardScaler, where each column is normalised to have a mean of 0 and a standard deviation of 1. An instance of the classifier model is then created with the number of hidden layers, neurons, and maximum iterations set. There are many ways to choose these numbers, but in the interest of simplicity, three layers are chosen with the same number of neurons as there are features in the data. The maximum iterations is a number that requires adjustment with the size of the data file, and through initial testing, 9000 is found to be an appropriate number for this implementation. With the data preprocessed and the model instance created, the classifier can begin to be trained and tested on the data.

This location data CSV file, with output genre labels attached to each line, is separated in an 80/20 split into training and testing data for the classifier. The training data is sent through the classifier, where all the inputs and the associated output are visible to the classifier, and weightings assigned to the inputs based on this data. The testing data is then given to the classifier, where again all the input data are available, but the output label is hidden. The classifier can then make its predictions based on the weightings previously set, and compare these predictions with the associated outputs, now made available. The results of this, combined with further testing will be discussed further in the *Evaluation* chapter of this dissertation.

5.3 Development & Implementation of Dynamic Pricing Model.

The final part of the implementation was to add a dynamic pricing model. This requires a method of computing the relevance of the predicted appropriate genre label to each of the stored adverts. As mentioned in the *Design* subsection of this chapter, the Advert class was initially designed to take in separate values for the input values, ie. time, season, weather, temperature, social class. With the newly included ANN Classifier giving outputs of advert genre labels, it became clear that this approach would be more appropriate in this implementation. Instead of taking in single input data values to pick adverts, the adverts were now initialised with a ranked ArrayList of relevant genre labels attached. Based on this ranking, the relevance of the predicted label can now be compared and used to adjust pricing as needed.

With research into billboard pricing and online advertising pricing, the value of €0.014 was seen to be a price applied to adverts being shown at off-peak times and was thus chosen as the base price in this implementation. This is the price that it will cost to show an advert that does not have the current appropriate label associated at all, the least relevant possible advert to show at the time. An advert with the predicted label as their first priority ranking in their ArrayList will cost €0.084 to show at this time. As the ranking of the predicted label decreases in this ArrayList, the cost of showing the advert at the given time also decreases, allowing the pricing to reflect the predicted effectiveness and appropriateness of the advert.

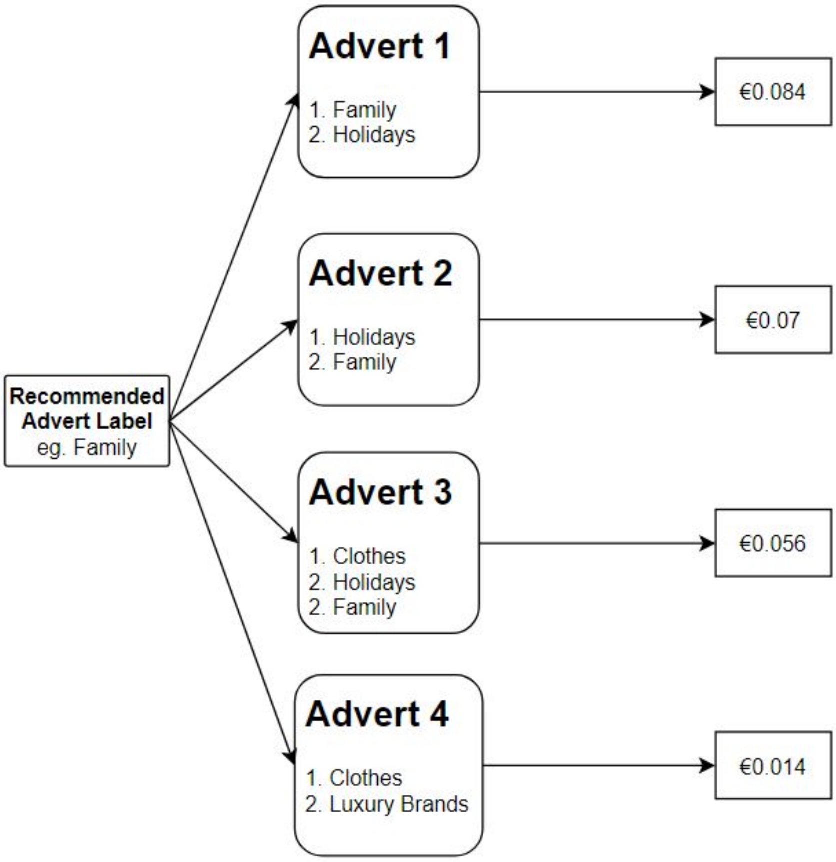


Figure 5.6: Advert pricing with sample predicted label of "Family"

5.4 Scalability

An important element of this implementation is the possibility of expanding and adding to its usefulness and effectiveness. The main method by which this implementation can be scaled up is with the addition of data streams and advert genre labels. This section will discuss how this scaling is made possible in the design of this implementation.

The addition of a data stream, depending on the method of retrieving this data, would involve the inclusion of a function within the data program to retrieve the values. This could be achieved through CSV analysis, Java library calls, API calls, or another method. To use this data in the training data for the ANN Classifier however, this data must be in integer form. This may require data conversion, or the inclusion of a form of classification of the ranges of values the data can result in (for example, in Table 4.2 the temperature values are put into classes, and each of these classes is given a number value ie. 1 for extremely freezing, 2 for freezing etc. for input into the CSV file). Adding an extra column to the training CSV file will require the data name to be added to the Python file for ANN training as it reads in the values from the data file, but that is the only necessary Python edit. Within the training data itself, the developer must determine how this new data stream will affect the output genre label. Then, these decisions must be included in the training data so that the classifier can assign weightings to this input and learn how it affects the output.

Additional advert genre labels are added to the system in a similarly uncomplicated way. Again, its addition into the training data CSV for the ANN classifier is an integral step in the process. The effect of the given input data streams on the new label would need to be included in the data file so that again, the classifier can be trained with the data and learn how to predict the new advert label as an output. The inclusion of a new genre label does necessitate some edits in the dynamic pricing model. The current pricing model allows for adverts to have up to five genre labels to be associated. It also provides the possibility of no genre labels being attached. With the new label it would be necessary to add in a price in the unlikely event that an advert has all possible genre labels associated, which is now maximum six as opposed to the previous five, and so on as further labels are integrated.

With the current implementation, scalability is completely doable and relatively uncomplicated, and is unquestionably a necessary aspect of the dissertation. This implementation is designed to be developed and improved upon, to increase its effectiveness of personalisation and targeting, and its overall usefulness.

6 Evaluation

The evaluation of the implementation of this dissertation was split between the neural network classifier and the dynamic pricing model. The testing of the classifier was built into the Python script used to create the model, and JUnit test cases were created to test the pricing model. These will be explained further in this chapter.

6.1 Neural Network Classifier

The first step of testing for the neural network classifier was to generate a confusion matrix after training the classifier, and running the test data. A confusion matrix, also known as an error matrix, is a visual representation of the model's performance, typically used in Machine Learning. The columns represent the possible outputs of the algorithm , while the rows show the predicted outputs that were generated - Predicted vs True values. The correct predictions are shown on the table in a diagonal line from the top left corner, to the bottom right, with any errors being outliers from this diagonal line.

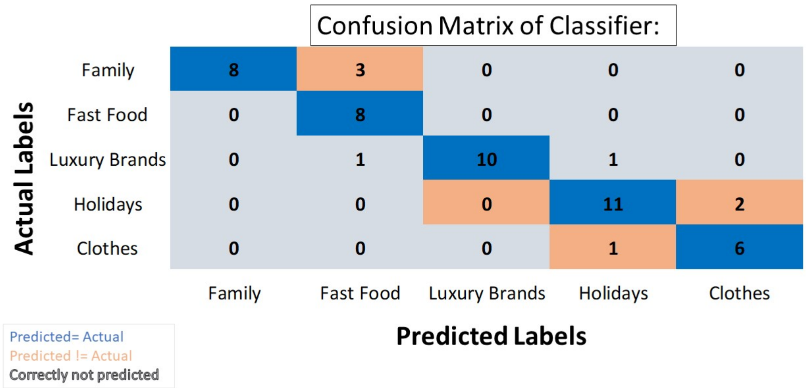
The second part of evaluation for the ANN classifier was to determine the precision, recall and F1-score. These are values used commonly within Machine Learning and classification to measure accuracy and reliability. "Precision" is defined as the number of true positives divided by the number of true positives plus the number of false positives - true positives being instances correctly classed as positive and false positives being cases where the model has incorrectly labelled an instance as positive when it should be negative. "Recall", in comparison, divides the number of true positives over the number of true positives plus the number of false negatives - false negatives being instances that the model labelled as negative that should have been positive. Precision is also known as the Positive Predictive Value and answers the question of how many positive classifications were correct. Recall is also referred to as Sensitivity or the True Positive Rate, and measures the proportion of positives that are correctly identified. The core difference being that recall is the ability to identify *all* possible positive instances, whereas precision identifies *only* relevant positive instances. For example, in a situation where a model is attempting to predict whether a plane passenger is a flight risk, the recall would be at 100% if all passengers were marked as flight risks, as no false negatives would be possible. On the opposite side, if all passengers are marked as not being flight risks, there will be a very high precision rate, as there would be no false positives (as statistically there are not many passengers who prove to be flight risks). The extremes of either of these measurements are not realistic and a balance must be found between them so they are useful in their context.

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \qquad recall = \frac{true\ positives}{true\ positives\ + false\ negatives}$$

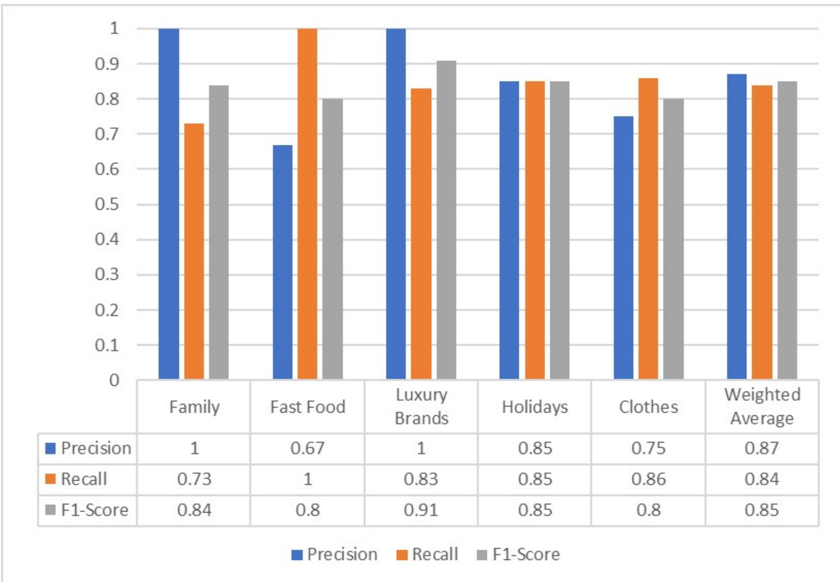
As we increase the recall value of a model, we decrease the precision, and vice-versa. In some cases it is clear which value should be focused upon, but generally a combination of both is needed, and in such instances the F1-score is typically used. The F1-score is the harmonic mean of precision and recall and is used over the simple mean as it gives equal weight to both precision and recall. This means that a model with 100% precision but 0% recall would get an F1-score of 0 as opposed to the 0.5 that a simple mean would give.

$$\textbf{F1 Score} = 2 * \frac{\text{Precision * Recall}}{\text{Precision + Recall}}$$

Testing was done on different configurations of the ANN Classifier model, particularly the number of nodes in the hidden layers and the number of iterations, as well as adjusting the number of hidden layers.

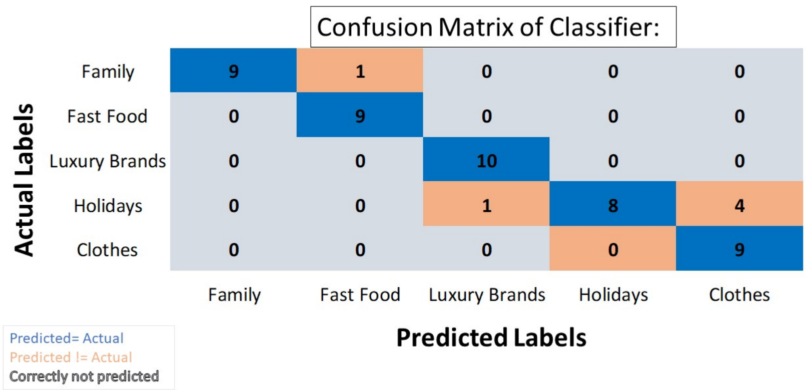


(a) Confusion Matrix.

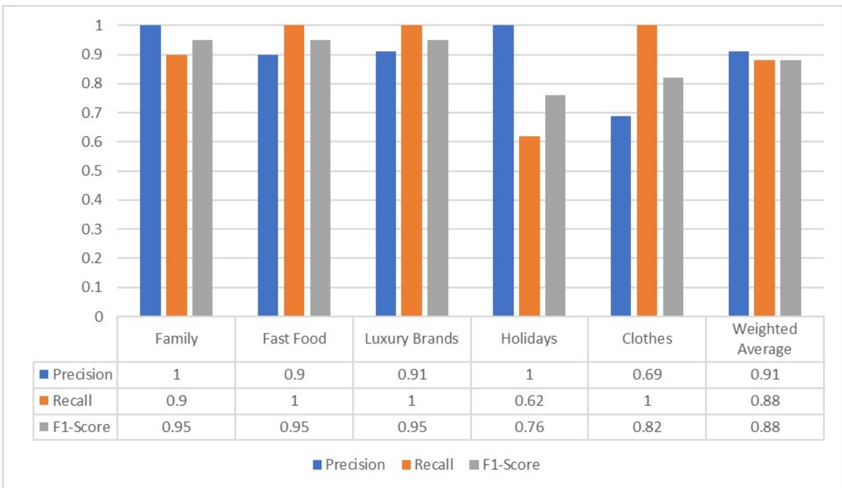


(b) Precision, Recall, F1-Score.

Figure 6.1: Model with 5 nodes in 3 hidden layers with 8000 iterations.

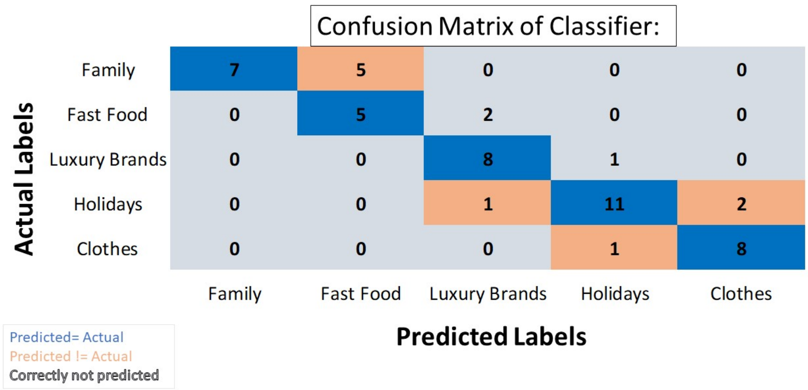


(a) Confusion Matrix.

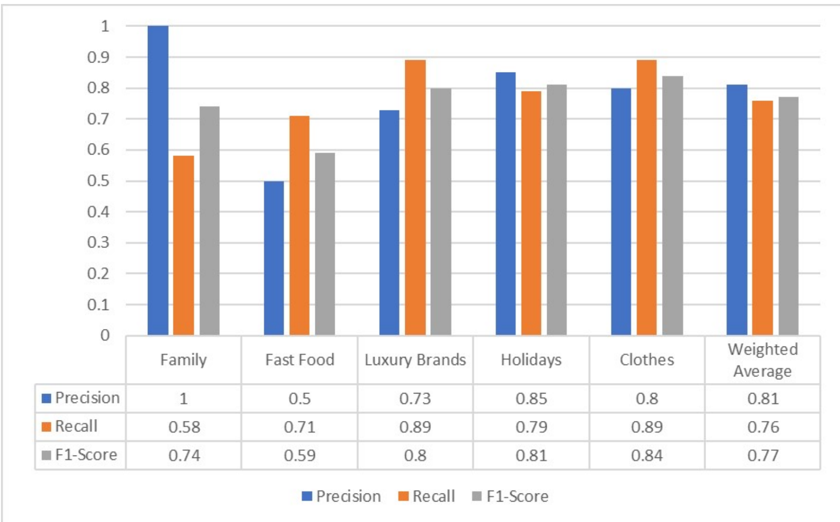


(b) Precision, Recall, F1-Score.

Figure 6.2: Model with 6 nodes in 3 hidden layer with 8000 iterations.



(a) Confusion Matrix.



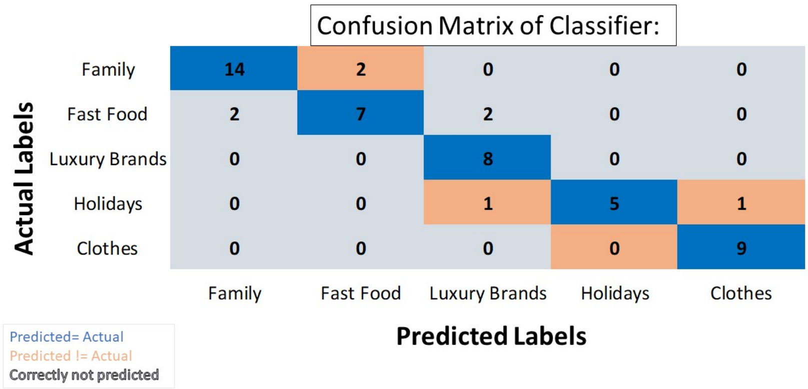
(b) Precision, Recall, F1-Score.

Figure 6.3: Model with 7 nodes in 3 hidden layer with 8000 iterations.

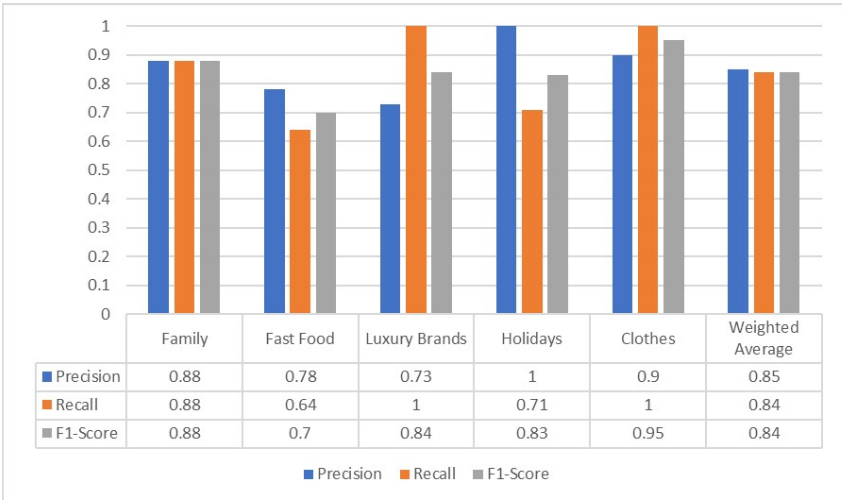
From this testing of varying the number of nodes in the hidden layers, a configuration featuring 6 nodes is shown to return the highest F1-score when compared to 5 and 7 nodes with 8000 iterations. As shown in Table 5.1, a configuratio of 6 nodes returns the highest F1-Score result. The testing continues, but having accepted 6 nodes as the apparent optimum, the number of iterations is varied to see the impact of change, and which value runs best with 6 nodes in the hidden layers.

Table 6.1: Node Number and F1-Score in model with 3 hidden layers and 8000 iterations.

Node #	F1-Score
5	0.85
6	0.88
7	0.77

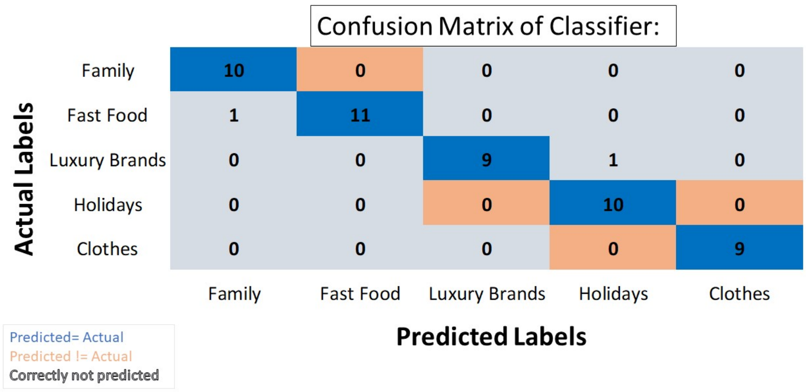


(a) Confusion Matrix.

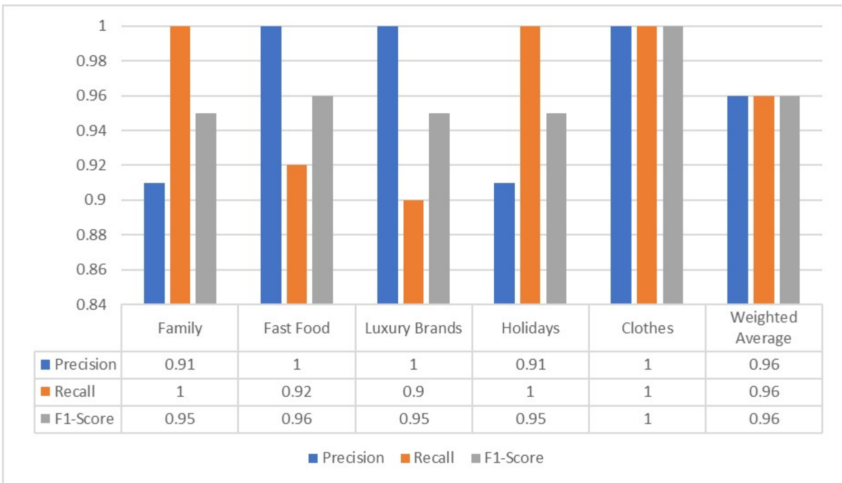


(b) Precision, Recall, F1-Score.

Figure 6.4: Model with 6 nodes in 3 hidden layer with 7000 iterations.

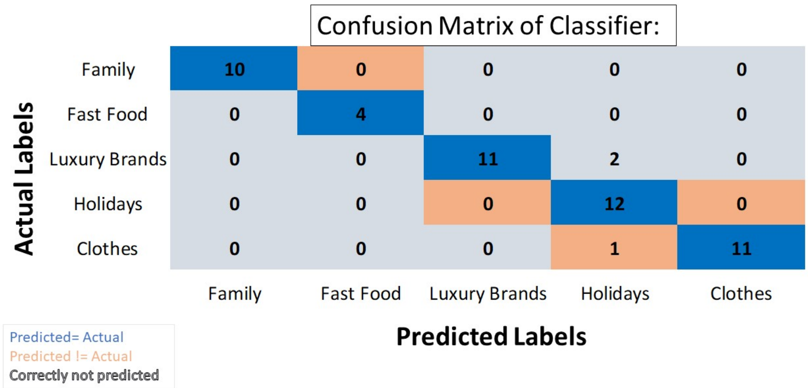


(a) Confusion Matrix.

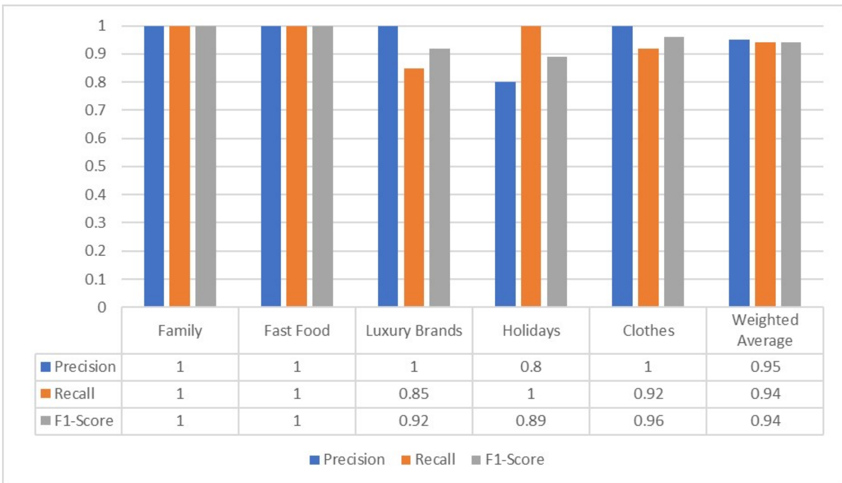


(b) Precision, Recall, F1-Score.

Figure 6.5: Model with 6 nodes in 3 hidden layers with 9000 iterations.



(a) Confusion Matrix.



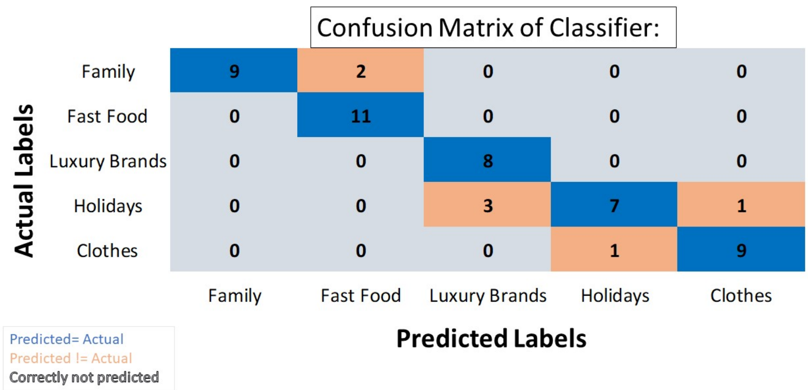
(b) Precision, Recall, F1-Score.

Figure 6.6: Model with 6 nodes in 3 hidden layers with 10000 iterations.

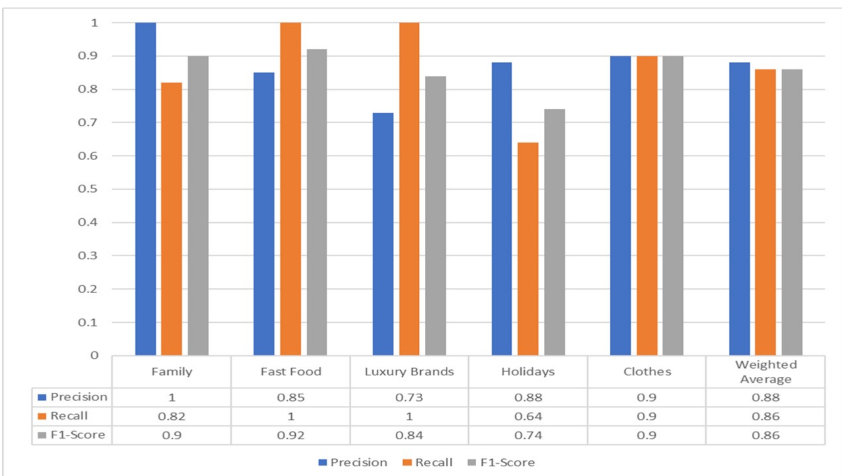
From this testing of iteration variation, 9000 iterations is the optimal number of iterations to use in conjunction with 6 nodes per hidden layer. The final element to adjust and test is the number of hidden layers in the ANN classifier while having 6 nodes per layer, and 9000 iterations.

Table 6.2: Iteration Number and F1-Score in model with 3 hidden layers and 6 nodes per hidden layer.

Iterations #	F1-Score
8000	0.88
7000	0.84
9000	0.96
10000	0.94

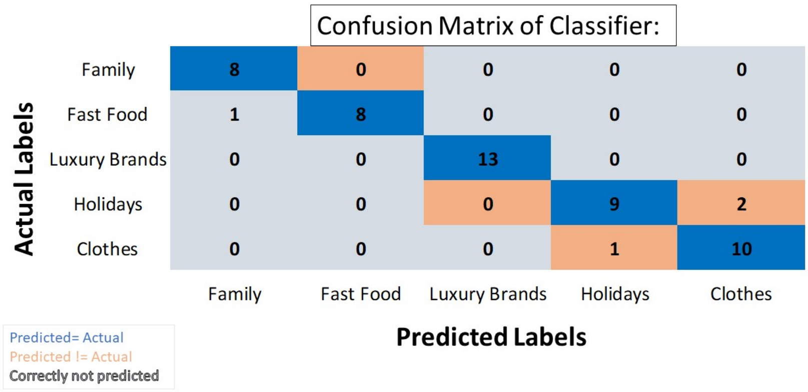


(a) Confusion Matrix.

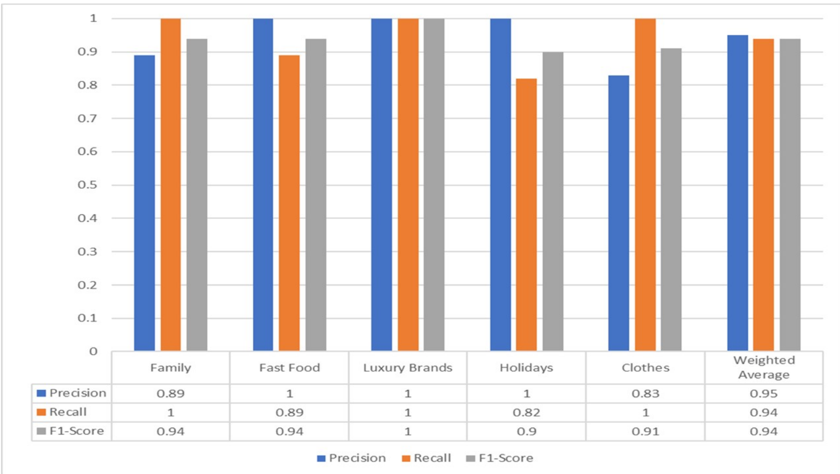


(b) Precision, Recall, F1-Score.

Figure 6.7: Model with 6 nodes in 2 hidden layers with 9000 iterations.

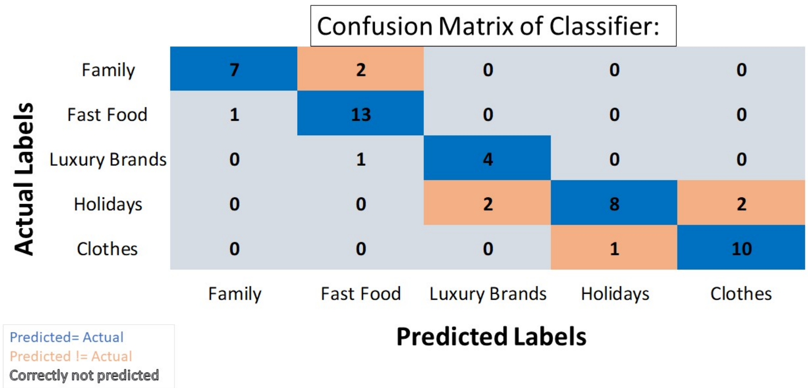


(a) Confusion Matrix.

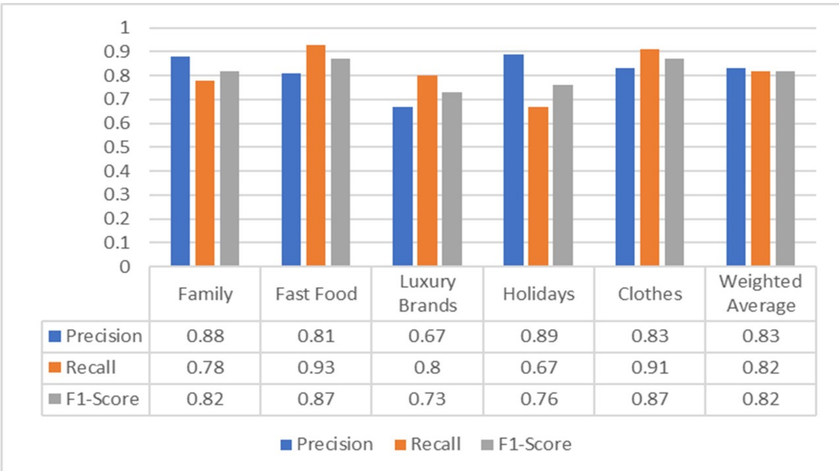


(b) Precision, Recall, F1-Score.

Figure 6.8: Model with 6 nodes in 4 hidden layers with 9000 iterations.



(a) Confusion Matrix.



(b) Precision, Recall, F1-Score.

Figure 6.9: Model with 6 nodes in 5 hidden layers with 9000 iterations.

The impact of altering the number of hidden layers is demonstrated by this final aspect of testing of the Neural Network Classifier model. In combination with 6 nodes per layer, and 9000 iterations, 3 hidden layers is seen to be the optimal number to implement (as seen in Figure 5.3). This will be the configuration used going forward with this dissertation, as the average F1-score is 0.96, the highest found through the testing, and most promising for its purpose as a classifier.

Table 6.3: Hidden Layer number and F1-Score in model with 6 nodes per hidden layer and 9000 iterations.

Hidden Layer #	F1-Score
2	0.86
3	0.96
4	0.94
5	0.82

6.2 Dynamic Pricing Model

Testing of the dynamic pricing model was completed through JUnit testing. Adverts were created with their associated ranked ArrayLists of genre labels. For this testing, six adverts were included with different permutations of the possible genre labels.

Figure 6.10: Adverts with genre ArrayLists for testing

By inputting testing predicted genre labels, the accurate output can be set and known. From this we can then use the *assertEquals* function within Junit testing to ensure that the outputs given from the program are the correct values.

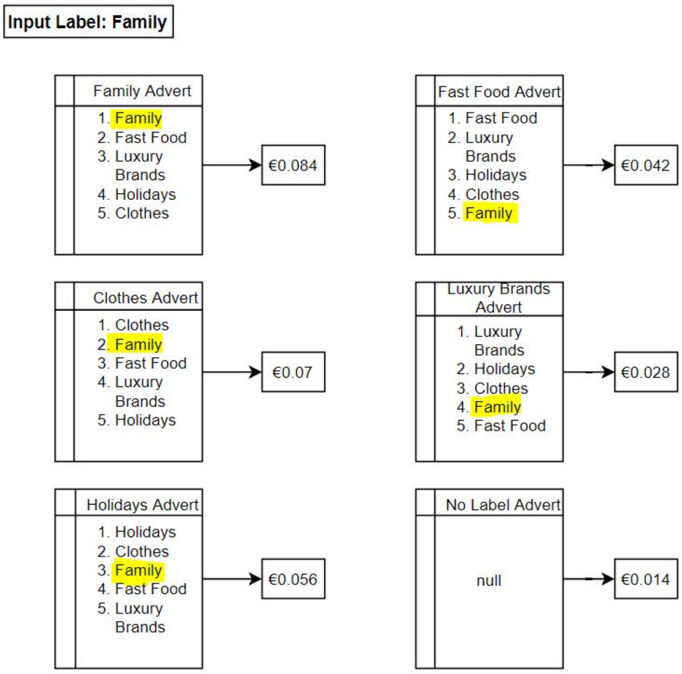


Figure 6.11: Advert pricing with Family label input.

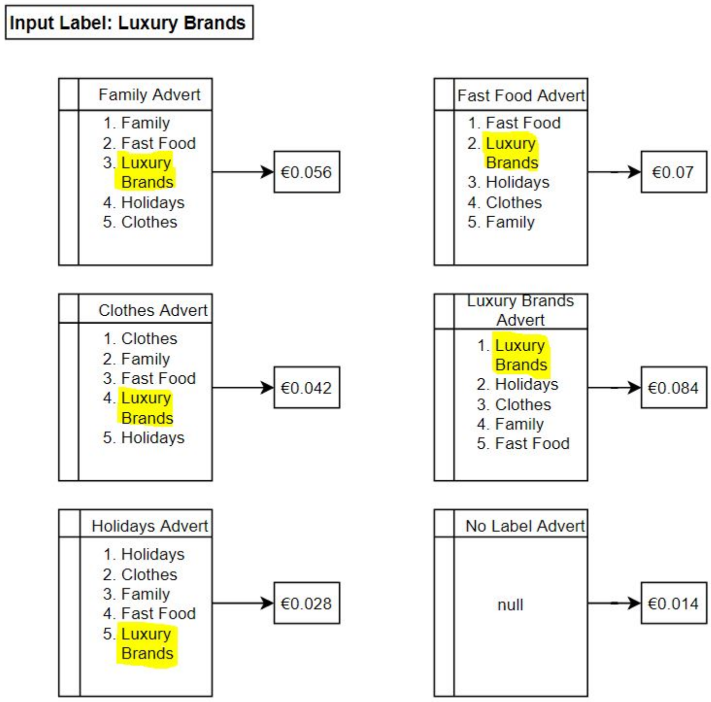


Figure 6.12: Advert pricing with Luxury Brands label input.

This testing was run with every possible input genre and, with some adjustment needed, was completed with zero errors. With the dynamic pricing model evaluated to be accurate, evaluation of the components of the dissertation was complete.

7 Conclusion

7.1 Findings

This dissertation proposed an implementation of a system that would include the benefits of OOH advertising (eg. constant visibility, broad audience reach), with the benefits of online advertising (eg. personalisation, targeting), as well as a concurrent dynamic pricing model. In the *Evaluation* chapter, it was demonstrated that this system was operable, and could return promising results. The location-based, and simulated, data, was collected and the ANN classifier was implemented on this to an average accuracy of 96%. Further, the dynamic pricing model successfully passed the JUnit testing fully, indicating that the components of this system work well together and offer a promising avenue for further research.

7.2 Reflection

The importance of research into decision-making methods led to some initial delays in the outset of this dissertation. This research was critical as the decision-making protocol was an integral part of this implementation. It also led to the discovery of research that emphatically supported the use of ANNs to improve marketing campaign effectiveness (15). A steep learning curve was experienced in the area of Machine Learning and these Artificial Neural Networks. Research and trial and error was required in finding the correct type of ANN as well as the configuration once the classifier type was found and implemented.

Within the decision-making protocol and training data of the classifier, there is a necessity to decide which output advert genre labels match which inputs. As advertising and data analysis companies do not release this data based on their research, this mandated assumptions be made based on research papers and public information. While this was enough to prove the effectiveness and usability of the implementation, this is a critically important aspect of advertising and more solid data and research, with fewer assumptions being required, would greatly benefit this component of the dissertation.

The initial outline of this dissertation included a broader range of location-based data streams, particularly nearby events, footfall, and nearby places of interest. Significant research went into determining the most appropriate method of calculating the number of people walking by the billboard. Some methods included a smartphone-based counter using Wi-Fi signal data (31, 32), or using video sensors (33, 34). The most promising option, however, in this implementation, was a Panasonic reflective photoelectric sensor that would count the number of times people passed through a given area in front of the sensor. Unfortunately, due to a lack of contact from the suppliers and a tight schedule, this aspect of the location-based data had to be discarded in this implementation.

In the instances of collecting information on nearby events and places of interest, there issues presented with the level of specificity allowed within the API calls to the event-related APIs (Ticketmaster and PredictHQ), and the nearby places of interest API (Google Places). For example, the Ticketmaster API returns all events on the website at the current time, without a method of filtering to only return or show events that are actually currently occurring, or occurring within a time frame. This resulted in volumes of irrelevant data being returned of events that could be booked, sometimes months, in advance, and was not useful in gauging the type of audience that would be in the area at the current time. Issues also arose with Google Places in that a number of the useful functions and specific API calls that had been shown during the initial research phase of the dissertation, have been deprecated since the papers had been written. Similarly, these API calls resulted in a large amount of unnecessary and irrelevant data being returned and needing to be filtered. The time-consuming nature of these data-streams with these APIs, as well as the questionable relevance of the actual data being returned, led to these APIs and data streams being discarded in the final implementation.

7.3 Future Work

Additional data streams in this implementation would be of benefit and could be added with more time and research. This could include footfall, nearby events and nearby places of interest, which were unfortunately unable to be added to this implementation. The increase in location-based data allows for more detailed and accurate personalisation of the area. In short, the more that is known of the area, the more effective the personalisation and targeting of adverts can be.

As well as additional input data streams, further analysis of possible advert genre labels as the output of the predictive model would broaden the usefulness of the implementation. There are currently the five outputs of family, fast food, luxury brands, holidays and clothes, but these could be both expanded on, with sub-genres (eg. fast food with the sub genres of soft drinks or dine-in vs. takeaway meals, or holidays with the sub genres of Winter or Summer holidays, or national/international), or added to (eg. a genre for films or upcoming events).

In the future, it would also be hugely beneficial to have more research done into the connection between the data inputs and the advert genres that are effective with the changes in data values. This could be done as a study or individual research, or potentially as a mutually beneficial partnership with an advertising or data analytics company. With a goal of decreasing assumptions made in the implementation, particularly the training data for the predictive model, this area of future development is potentially of the most importance, as the predictive and personalisation aspect of the implementation is at the core of its effectiveness.

Bibliography

[1] Nielsen Norman Group Jakob Nielsen. F-shaped pattern for reading web content. URL https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/.

[2] Keno Leon. Making a simple neural network, . URL https://becominghuman.ai/making-a-simple-neural-network-2ea1de81ec20.

[3] Glosser.ca. Artificial neural network.svg, 2013. URL https://commons.wikimedia.org/w/index.php?curid=24913461.

[4] K. O'Donnell and H. Cramer. *People's Perceptions of Personalized Ads*. 2015. WWW '15 24th International World Wide Web Conference.

[5] ATA Outdoor. Top 10 reasons to consider outdoor advertising. URL https://www.ataoutdoormedia.com/outdoor-advertising-top-10/.

[6] J. David Lichtenthal, Vivek Yadav, and Naveen Donthu. Outdoor advertising for business markets. *Industrial Marketing Management*, 35("2"):236 – 247, 2006. doi: https://doi.org/10.1016/j.indmarman.2005.02.006.

[7] The Manifest. How businesses invest in digital marketing in 2018. URL https://themanifest.com/digital-marketing/how-businesses-invest-digital-marketing.

[8] ARK Analyst James Wang. How internet advertising can grow to $600 billion by 2023. URL https://ark-invest.com/research/internet-advertising-growth.

[9] Jivox. 2016 jivox benchmark report, 2016.

[10] Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Trans. Comput.-Hum. Interact.*, 12(4):423–445, December 2005. ISSN 1073-0516. doi: 10.1145/1121112.1121116. URL http://doi.acm.org/10.1145/1121112.1121116.

[11] Nielsen. Poster advertising study, 2017.

[12] Juntao Lai, Tao Cheng, and Guy Lansley. Improved targeted outdoor advertising based on geotagged social media data. *Annals of GIS*, 23(4):237–250, 06 2017.

[13] Akhilesh Kumar Sinha and Neeraj Kumar. Demand response managemengt of smart grids using dynamic pricing, 2016. 2016 International Conference on Inventive Computation Technologies (ICICT).

[14] Yuan Chen and Feifei Wang. A dynamic pricing model for e-commerce based on data mining. 2009 Second International Symposium on Computational Intelligence and Design.

[15] Gianluigi Guido, M Prete, Stefano Miraglia, and Irma De Mare. Targeting direct marketing campaigns by neural networks. *Journal of Marketing Management*, 27: 992–1006, 08 2011. doi: 10.1080/0267257X.2010.543018.

[16] Badhe Anup. Using deep learning neural networks to find best performing audience segments. *International Journal of Scientific & Technology Research*, 4.8:30–31, 2015.

[17] Keno Leon. Making a simple neural network : Classification, . URL https://becominghuman.ai/making-a-simple-neural-network-classification-2449da88c77e.

[18] Kurt Iveson. Branded cities: outdoor advertising, urban governance, and the outdoor media landscape. *Antipode*, 44(1):151–174, 1 2012. ISSN 1467-8330. doi:

10.1111/j.1467-8330.2011.00849.x. URL
https://doi.org/10.1111/j.1467-8330.2011.00849.x.

[19] Keith Kelsen. 1 - the new medium: Digital signage and the power of content. pages 1 – 12, 2010. doi: https://doi.org/10.1016/B978-0-240-81302-8.00001-4. URL http://www.sciencedirect.com/science/article/pii/B9780240813028000014.

[20] Daniele Quercia, Rossano Schifanella, Luca Maria Aiello, and Kate McLean. Smelly maps: The digital life of urban smellscapes. *CoRR*, abs/1505.06851, 2015. URL http://arxiv.org/abs/1505.06851.

[21] Rossano Schifanella, Daniele Quercia, Francesco Aletta, and Luca Aiello. Chatty maps: Constructing sound maps of urban areas from social media data. *Royal Society Open Science*, 3, 03 2016. doi: 10.1098/rsos.150690.

[22] Paul A Longley, Muhammad Adnan, and Guy Lansley. The geotemporal demographics of twitter usage. *Environment and Planning A: Economy and Space*, 47(2):465–484, 2015. doi: 10.1068/a130122p. URL https://doi.org/10.1068/a130122p.

[23] S. Nigam, S. Asthana, and P. Gupta. Iot based intelligent billboard using data mining. In *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, pages 107–110, Feb 2016. doi: 10.1109/ICICCS.2016.7542329.

[24] Hidayet Aksu, Leonardo Babun, Mauro Conti, Gabriele Tolomei, and A Selcuk Uluagac. Advertising in the iot era: Vision and challenges. *IEEE Communications Magazine*, (99):1–7, 2018.

[25] The Digital Signage Insider. Dooh ads sell for average cpm of $12: Survey. URL https://www.wirespring.com/dynamic_digital_signage_and_interactive_kiosks_journal/articles/DOOH_Ads_Sell_for_Average_CPM_of__12__Survey-809.html.

[26] Sebastian Gutierrez. How to make better use of time of day in targeted advertising. URL https://econsultancy.com/how-to-make-better-use-of-time-of-day-in-targeted-advertising/.

[27] Chloe He. Why timing is everything for your marketing campaigns. URL https://www.business2community.com/marketing/timing-everything-marketing-campaigns-0996721.

[28] Think Metric. Temperature with think metric. URL https://thinkmetric.org.uk/basics/temperature/.

[29] Central Statistics Office. Census of population 2016 – profile 9 health, disability and carers, appendix 2 - definitions, 2016. URL https://www.cso.ie/en/releasesandpublications/ep/p-cp9hdc/p8hdc/p9bgn/.

[30] Lukar E. Thornton Philippa J. Settle, Adrian J. Cameron. Socioeconomic differences in outdoor food advertising at public transit stops across melbourne suburbs. 10 2014. Australian and New Zealand Journal of Public Health.

[31] Haochao Li, Eddie C. L. Chan, Xiaonan Guo, Jiang Xiao, Kaishun Wu, and Lionel M. Ni. *Wi-Counter: Smartphone-Based People Counter Using Crowdsourced Wi-Fi Signal Data*, volume 45. 2015. IEEE Transactions on Human-Machine Systems.

[32] N. Ahmed, A. Ghose, A. K. Agrawal, C. Bhaumik, V. Chandel, and A. Kumar. Smartevactrak: A people counting and coarse-level localization solution for efficient evacuation of large buildings. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 372–377, March 2015.

[33] M. Irfan, L. Marcenaro, and L. Tokarchuk. Crowd analysis using visual and non-visual sensors, a survey. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1249–1254, Dec 2016.

[34] Shizuka Fujisawa, Go Hasegawa, Yoshiaki Taniguchi, and Hirotaka Nakano. Pedestrian counting in video sequences based on optical flow clustering. *International Journal of Image Processing*, 7(1):1–16, 2013.

A1 Appendix

A1.1 Security Considerations

The title of the dissertation in question is "Real-Time, Targeted, Out-Of-Home Advertising with Dynamic Pricing". The implementation consists of:

1. Data collection
2. Training of a Neural Network Classifier on this data
3. Setting price of adverts based on Classifier output
4. Presenting relevant advert to a digital billboard

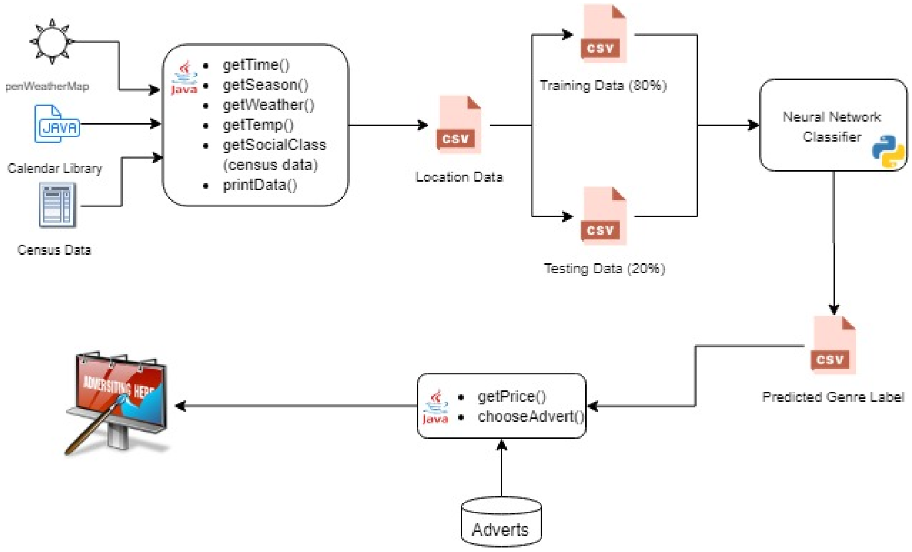


Figure A1.1: Implementation Architecture

The data collected for this dissertation include the current weather and temperature, collected through API calls to OpenWeatherMap, the time and season, determined by the

Java Calendar library, and the location's social class, using census data from the Central Statistics Office, presented as CSV files.
The dependency on the OpenWeatherMap API interface opens up the project to potential issues if OpenWeatherMap has vulnerabilities that could be exploited in the future. The connection to this implementation also opens up their API to potential attacks, if the code of this project was altered to send problematic code to the API. Unless OpenWeatherMap have suitable security measures in place, their data could be affected. The use of API injections could be used, which is when malicious code is inserted into an API call. If the API doesn't have suitable measures in place, such as a Filter Input, Escape Output (FIEO), the malicious code could initiate a cross site scripting attack targeting the end-users browsers, or an SQL injection where the databases of the servers are affected. FIEO is the process of filtering the input and preventing harmful known characters or commands from being performed, and escape output is the method of preventing an API call from performing an output that is not supported by the API design, and this is one method of attempting to prevent injection attacks. This injection prevention method could also be implemented on the side of the implementation code as a layer of protection for the OpenWeatherMap API to attempt to prevent any issues if the API calls are altered. This Java program of data collection is run on a central computer in this implementation so code alteration would involve the attacker to have physical control of the computer running the code. In the future this code would potentially be run on a Raspberry Pi connected to a digital billboard, which would negate the need for access to the central computer.
As well as API injection attacks, through the API calls in this project, an attacker could initiate a Denial of Service (DoS) attack. This is when the number of API calls is hugely increased to attempt to overload and overwhelm the application or API, causing disruptions or even complete stoppages of the hosts services. To prevent this hosts can implement traffic rate controls where the number of requests cannot be greater than a certain amount or network control will be initiated and malicious IP addresses blacklisted.
Though not fully implemented within the time constraints of this dissertation, in the future there is a necessity for access to be allowed for the users of this implementation, in this

context being advertisers. The interactions of these users with the system would be logging into an associated website to upload their adverts and then choose a pricing model based on the dynamic pricing of the project. There would be a login requirement so that users would have their separate data and not be able to access other users' data, but this also opens up the system to potential web attacks. Injection attacks are again an issue, with SQL injections being the number one problem on the list of Top 10 Security Issues put out by the Open Web Application Security Project (OWASP). This is when an attacker inputs an SQL command into the site, typically through the login form or the url of the site itself, to attempt to force the database to perform the command, be it data alteration, deletion or a request to view the data. As mentioned before, filtering of user input is a large component of preventing injection attacks. Each user input section should only allow relevant characters, for example, a section for phone numbers should only allow numbers be entered and no other types of characters. Appropriate privileges should also be implemented. If the user hasn't proven themself to be of the position to be able to access sensitive data, the database containing this data should not be connected to the user interface at all. User logins should only connect them to the data that is directly associated to them, no other user's data should be connected, to prevent a malicious user from accessing other people's data, and a breach at this level would only reveal their own data to them. Firewalls can also be used to filter out malicious code, though require consistent updates and patches to ensure newly found vulnerabilities are fixed.
In the current implementation of the dissertation, there is not any sensitive data saved so security issues will mainly arise in future work as user details, such as usernames, passwords and even payment details, are saved. Currently the API calls being made to OpenWeatherMap is the main connection of concern as everything else is contained in the system which would require physical access to the PC running the code to be altered. In this way the system seems quite safe and the lack of sensitive data reduces the temptation of hacking or attacking, but in the future as end-users are included, the above mentioned measures would have to be taken and adjustments made to prevent malicious attacks on user data.