**CE101 Team Report Assignment**

**Team Number- A14**

**Team Members - Brodie Pestell, Julian Plein, Plamen Petkov, Victoria Quest-Travis, Hikmat Ali, Syed Mohsin Raza and Trina Roy**

### CSEE Jira Project URL: <https://cseejira.essex.ac.uk/projects/CE101T14/summary>

**Table of Contents**

**Chapter 1 The Executive Summary (216 words) 3**

**Chapter 2 Team Working (1860 words) 3**

2.I An introduction to Team Working 3

2.II Team Activity Report 4

*2.II.a Team's Scrum Meetings, Sprints 4*

*2.II.b Detailed report of each team members contribution to the project 7*

**Chapter 3 Product Development (3044 words) 10**

3.I An introduction to Product Development 10

3.II The Context 10

*3.II.a The Customer/ End User 12*

*3.II.b Legal and Ethical Matters 12*

*3.II.c Sustainability 12*

*3.II.d Health and Safety Matters 13*

3.III The Team Product 15

*3.III.a Product Description 15*

*3.III.b Product Demonstration 17*

**Chapter 4: Project Management (?? words) 18**

4.I An introduction to Project Management 18

4.II Project Management Report 18

*4.II.a Summary of Teams Project management in Jira 18*

*4.II.b An evaluation of the project management 19*

**Chapter 5: Conclusions (334 words) 20**

# Chapter 1 The Executive Summary (216 words)

Our teams project was a little more on the difficult side. This is due to the fact that a few have only just started programming this university year and yet were tasked with machine learning. Despite this, we managed to work well, however due to the nature of this project, and the fact that this project is mostly code, it was difficult to have specialists, however, everyone was more delegated a “topical” role. Trina and Plamen’s role mostly consisted of research and had proven to be very strong, producing easy to understand result and always doing it on time. Victoria’s role was a sort of mixed role which was an optimal role as she had some coding experience but nothing near as complex as this. Brodie and Julian were the main delegated programmers as they have had the most experience and had the highest understanding of neural networks. Mohsin also had limited prior programming so contributed mainly to research and did some coding which wasn’t used in the end. Some roles would cross over, however they mostly remained consistent. Our team had a strong start as we were completing tasks online and most people were able to attend all sessions. Unfortunately, we did not have weekly meetings due to schedule clashes and timetabling issues. Doing so could have ensured that we finished quicker and therefore, would allow us to make more adjustments to higher our Kaggle rank. A key trait that we all picked up was discipline where it required that we all manage and prioritise different stages of product development. Working to everyone's strengths and weaknesses was also a very rewarding experience and it was an interesting part of the learning curve. Saying that, the overall experience was new and refreshing and allowed everyone to reinforce as well as pick up new skills along the way which we are sure can be implemented directly into everyday life.

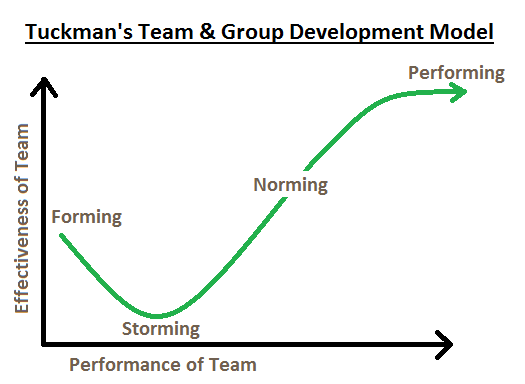
# Chapter 2 Team Working (1860 words)

## 2.I An introduction to Team Working

Teamwork is an important skill that every one of us has. It is about forming a team with people from different backgrounds, knowledge level and experience to collaboratively and efficiently work towards achieving a common goal. [1]

Working as a team has its advantages and disadvantages. The main advantage is that the project can be split into separate tasks which then can be grouped under different skill requirements. This allows to assign individual tasks suited to each member’s skills and evenly distribute the work among all members. However, in order to achieve this, team members have to trust each other that they will complete their assigned task. [1][2]

There are two team formation models that stand out above the rest. One of them is Bruce Tuckman’s model: Forming, Storming, Norming and Performing (later on an Adjourning stage is added), and the other one is George Charrier’s: Cog’s Ladder. They are both very similar in nature but Tuckman’s qqmodel is more widely known. As the name suggests, Tuckman’s model is made out of four stages (the fifth stage is added twelve years after the creation of the model), the model starts with the Forming stage where the team members meet each other and learn about their task. The second stage is Storming. At this stage, team members discuss various ideas on how to go about the task. Conflicts are likely to arise. The actual development of the product or task begins at the third stage- Norming, where the team starts discussing ideas and setting rules. On the final stage - Performing, the team implement their solutions and ideas.

The graph above represents the effectiveness of the team during different stages of Tuckman’s model.[3]

The next step after forming a team is assigning roles to each member. Dr Meredith Belbin breaks the team roles into a total of nine behavioral team roles with which team members can align to more than one. The nine roles are split into three groups:[4][5]

1. Action Oriented Roles
   1. Shaper - A person who always pushes the team forward by giving them new challenges and exploring new ideas.
   2. Implementer - A person who is able to take an idea and turn it into reality.
   3. Complete Finisher - A person who ensures that the project is completed and there are not any errors left.
2. People Oriented Roles
   1. Chairman - A person who is able to guide the team as a Team Leader.
   2. Team Worker - A person who supports other member’s and makes sure that others work effectively.
   3. Resource Investigator - A person who provides ideas, resources and information for the team.
3. Thought Oriented Roles
   1. Plant - A person who is creative and imaginative. He comes up with the ideas but ignores the details.
   2. Monitor-Evaluator - A person who is sober and helps keeping the team ‘on track’.
   3. Specialist - A person who passionately brings specialist knowledge to the team.

## 2.II Team Activity Report

## 2.II.a Team's Scrum Meetings, Sprints

We held a total of 9 meetings and had 5 different sprints regarding the project. Every sprint, we had a new scrum master who delegated roles and tasks as well as managed the group’s progress.

Sprints

Sprint 1:

Scrum Master: Julian Plein

Date: 9/11/18 – 18/11/18

• Learn about what data will be fed into the machine learning algorithm in the first place, to allow it to derive a price from the variables and what the process looks like. (Assigned to Victoria)

• Look into machine learning algorithms and be able to, upon completion, explain it to the rest of the group. Only a basic understanding is needed. (Assigned to Plamen)

• Look into how regression functions in python and be ready to explain it to the rest of the team. (Assigned to Trina)

• Retrieve and look over the code of a previous simple neural network. Share insight. (Assigned to Julian)

• A GitHub repository needs to be created and the team members need to be invited. (Assigned to Brodie)

• Catch Up –Please read the lab notes for unit 3 and 4 on Moodle to catch-up with work done this week and last. (Assigned to Mohsin)

Sprint 2:

Scrum Master: Victoria Quest-Travis

Date: 23/11/18 – 30/11/18

• Start histograms, scatter plot, box plot, bar chart and correlation. (Assigned to Trina)

• Determine whether a neural network is suitable to the task by adapting the previously retrieved network (or creating a similar one). (Assigned to Julian)

• Univariate and Multivariate-Adapt what you have done already to include multiple variable in the graphs. (Assigned to Plamen)

• Catch Up –Please read the lab notes for unit 3 and 4 on Moodle to catch-up with work done this week and last. (Assigned to Mohsin)

• Learn about different transformations and training machine learning models. Apply them both. (Assigned to Victoria)

• Understand how variables are distributed and how they interact. (Assigned to Brodie)

Sprint 3:

Scrum Master: Brodie Pestell

Date: 7/12/18 – 18/12/18 (date to be fixed)

• Feature selection- Learn about and implement feature selection using filter and wrapper. (Assigned to Brodie)

• Learn about, and how to use, data visualisation techniques to visualise distributed and missing data. Victoria has already been working on determining missing data, so work with her to determine which data is missing and then visualise it. (Assigned to Trina)

• Work with Trina to plot heatmaps and show correlations of qualitative and quantitative data. (Assigned to Victoria)

• Integrate loading the test/training data into the neural network for training with help from Victoria. (Assigned to Julian)

• Learn about and understand linear regression as a concept and in terms of the house prices. (Assigned to Plamen)

• Convert neural network to Python 3. (Assigned to Julian)

Sprint 4:

Scrum Master: Plamen Petkov

Date: 25/1/19 – 1/2/19

• Work with Trina to plot heatmaps and show correlations of qualitative and quantitative data. (Assigned to Victoria)

• Train the dataset. (Assigned to Brodie)

• Research on Neural Network libraries. (Assigned to Julian)

• Ensure that you have in-depth Knowledge of linear regression. (Assigned to Plamen)

• Work with Victoria to plot heatmaps and show correlations of qualitative and quantitative data. (Assigned to Trina)

Sprint 5:

Scrum Master: Syed Mohsin Raza

Date: 22/2/19 – 1/3/19

• Start researching chapter 2 of the report and write bullet points/draft of what could go in that section. (Assigned to Plamen and Victoria)

• Start researching chapter 3 of the report and write bullet points/draft of what could go in that section. (Assigned to Hikmat, Mohsin and Brodie)

• Start researching chapter 4 of the report and write bullet points/draft of what could go in that section. (Assigned to Trina and Julian)

We also held meetings outside of the sprint times in order for everyone to be up to date on our project and in order to work together more to get the project done as well as add in a few extra features, that we thought was necessary.

23/1/19:

This is where the conversion of the neural network had been finished and the network was being tested to see if the weightings were correct and the code would run

6/2/19:

Version 2 of the neural network had just been created and the feature selection had been added to it.

8/2/19:

We got our first house price predictions and uploaded it to Kaggle to get our rank. Hikmat had also just joined our team.

### 2.II.b Detailed report of each team members contribution to the project

Julian G. Plein:

He came up with the idea of using a neural network. He provided the base neural network and converted it from Java to Python 3, based upon an online example. He also explained the basics of how it worked and how we could connect it to the project at hand. Although it did not work as planned (due to various code issues with the older original build), this idea and initial build allowed for another build to be created, based on the original idea. While the first version focused on a multi-layer neural network coded from scratch, this new attempt took advantage of online libraries to provide more accurate and more efficient results. He also helped with the 2nd build of the neural network (working with Brodie) allowing for a more efficient and adaptable network that allowed us to get more accurate results.

Trina Roy:

She was mostly in charge of graphs and plotting them. This involved using heatmaps, pair plots and bar graphs to show the correlation between all the features within the required attributes (such as the year built, the area of the garage). Using the correlation coefficients, she was able to detect multicollinearity (The state of high intercorrelations within the variables [6]) (pls add more as I am limited in the area of knowledge)

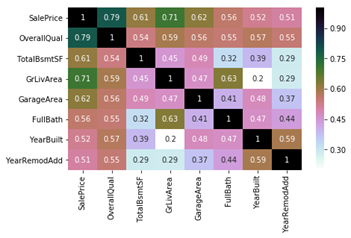


Chart 1: Heatmap showing correlation between the variables.

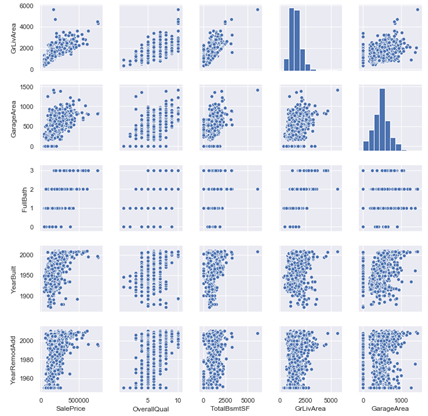
Brodie Pestell:

He was tasked with determining how the variables were distributed in the data and how they interacted with each other (how correlated they were to one another and to the house price). After this research was completed, Brodie was then made responsible for the feature selection of the data. This is heavily related to the initial task, and involves determining which features have the highest correlation to the house price. Feature selection is important as it prevents the model from being overtrained, which can reduce the accuracy of the model. Whilst developing the feature selection code, Brodie also helped Julian implement the neural network version of the code, which was originally designed by Julian in the past in a different language. After this, Brodie was put in charge of training the model on the dataset - this allows the model to actually make predictions, and is part of what affects the overall accuracy of the model.

Mohsin Raza:

He started off by doing some initial research on various topics but mainly on linear regression. He was fairly new to coding in general so his main contribution came through the form of research into linear regression and machine learning. He wrote a few pieces of code which were a potential input to the first draft however, as more knowledge was gained on how to make improvements, none of the code was used in the end. He was also a scrum master, delegating people roles in things such as the essay. In terms of the essay, he has contributed to parts: 1, 3.IIa, 3.IIb, 3.IIc and 3.IId as well as proof reading the essay for any spelling/grammatical mistakes and maintaining proper format throughout.

Victoria Quest- Travis:

Her main role was to figure out what data was missing and how to get that missing data and turn it into data that can be processed by our program. To do this, she used *GET DUMMIES* (which changes string values to number values in order for the machine to be able to use it), *FILL NA* (which changes all values from strings to numbers by creating a lot of temporary columns) and *DROP NA* (which drops all “NA” values but we did not use this one in the long run). She also tested out different transformations that could be done such as np.argsort to sort the variables in order of correlation (which was placed onto the heatmap, *.describe* to ( which finds of entries (rows), the mean, the maximum and minimum value as well as upper quartile and lower quartile if the values) .She also helped Trina with heatmaps and correlation and also pair plots (as seen below, not all are shown).

Plamen Petkov:

His main role was to research machine learning algorithms and find which one would be the best suited for our project, expand his knowledge of this and figure out how this was an important part to the machine learning code. He would then present to us the difference between Supervised machine learning algorithms where the learning model learns from the past input data and makes future predictions as an output, such as Classification (Decision Tree and Support Vector Machine) and Regression (Linear and Polynomial Regression), and Unsupervised machine learning algorithms where the learning model uses unlabeled input data and allows the algorithm to act on that information without guidance such as Clustering (K-means) and Association. This helped us find the right algorithm for our project.

Hikmat Ali:

I joined this team on 08/02/2019 due to having problems with my own team with people not showing up and people leaving university etc. When I joined the team they had just received their first house price prediction and uploaded it to kaggle. So the team already had done the planning, designing, code developing and was on to testing the code and seeing how much they scored on kaggle etc. So I couldn’t really have any input into this as I just joined this team and didn’t understand majority of what was happening. However, it is still my fault due to not contacting Vishu earlier about the problems I was facing. My contribution to the team was through the team report. I completed Chapter 3.1, 3.2.b with Mohsin, 3.2.d with Mohsin.

# Chapter 3 Product Development (3044 words)

## 3.I An introduction to Product Development

Product Development is the creation of products with new or different features that offer new or additional advantages to the customer. Product development could involve changing of a current product that is available or creating a completely new product that satisfies a mass or niche market.

There are several ways on how we can approach to develop a product (ad-hoc, hacking etc...) However, from the ways in which products can be developed we used the development methodologies [7]. There are various methodologies and choosing one before beginning a specified product is important for creating an efficient structure. Two of the more popular methodologies are the Waterfall (traditional) and Agile methodologies. However, many more exist: Crystal Methods, Dynamic Systems Development Model (DSDM), Extreme Programming (XP), Feature Driven Development (FDD), Joint Application Development (JAD), Lean Development (LD), Rapid Application Development (RAD) etc. [8]

The classic waterfall method, also known as the traditional method is a rigid linear model that consists of progressive phases. The methodology starts with the requirements of the project. This is followed by design stage, where the planning and designing of the project takes place. It also can take into aspects of the requirements of the project too. After this is done, the implementation stage starts. There could be feedback given to both previous stages on what can be changed before moving on too. Once the project has been implemented then it is tested to demonstrate it meets the specification. Also, maintenance can be done to the project too. [9]

Some advantages to the waterfall method are that the linear nature of the method makes it easy to understand and manage. The projects with clear cut objectives and steady requirements can best use the waterfall method. The method allows those who have less experience to benefit a lot from using the methodology [9]

Some disadvantages to the waterfall method are that the method can often be slow and can be costly due to the rigid structure and tight controls. Also, the solution that is provided at the end of the process may no longer be able to solve the current problem such that your solution is outdated. So due to these downsides it may lead users to explore other software development methodologies too. [9]

When deciding which methodology to pick for a project is largely reliant on how clear the initial requirements are and how clear the solution is from offset. An approach which is more dynamically flexible is the Agile methodology.

The Agile methodology is becoming more extensively used in software development and even in hardware development. It can adapt to solutions that it faces along the way of making the project. Therefore, it could be suitable for product development where the goals may be clear but the solutions very unclear. [10]

Each methodology has its own set of roles which can be accepted by team members. For instance, the Waterfall method generally has a project manager (ensures that the quality of the final product is good, manages projects and designates tasks between team members), a business analyst (whose job is for writing business strategy and ensures that the software product is popular in the market), a developer (designs the code), and a tester( who tests the final software product that is given and then feedbacks to the developer about what can improve.) [11]

Alternatively, The Agile methodology consist of three key roles within the Scrum framework; the Scrum Master (ensures the team is functioning well, helps the team by making sure they are working efficiently by removing any obstacles they face, obtaining resources for the team, creates meetings with the team and works with the product owner to maintain the product backlog.), the Scrum Team (they are responsible for completing the work agreed for the current sprint. They all work together to make sure to do modelling, programming, testing and release activities.) and the Product Owner (represents the vision of the project being built through a prioritised list (product backlog) and makes decisions in a timely manner.) [12]

Being creative in software development can be overlooked, yet it is vital as it is the source to solve complex problems and revolutionise [12]. The advocates of Agile software movement emphasize the importance of people, to the point that ‘People trump Process’ 12].

The pros of Agile includes attempt to minimize as much risk as possible when adding new functions by developing the software in iterations that are mini-additions of the new functionality. This benefit of doing this is the developer can see the faults in the development early on and can fix this. The cons of Agile methodology include reliance on real time communication, which fails to provide new users with documentation to get up-to-speed. You must have good communication skills with team members to make sure that each bit is completed and up to date on time.

## 3.II The Context

### 3.II.a The Customer/ End User

The whole aim of this project was to allow for the prediction of house prices, an area which is seen as being lacklustre so far by many in terms of accurate predictions. The sheer number of factors that need to be taken in make it a difficult but rewarding challenge to create and maintain a high-quality product. In relation to our product, we’d like to think that the target market would range anywhere from your average first time buyers all the way to those that have their feet sunk in deep into the housing business market. For people that are looking to offload their house, they’d be able to use it to realise the true potential value of their sale by looking at what their home would be worth down the line, enabling maximum profits. One of our aims was to create something unique and user friendly which would hopefully inspire others to innovate and create also. Creating a product like this for a large scale company can be deemed safe enough to make it profitable, since houses are always in demand. Therefore this correlates with constantly changing house prices. A potential loss of jobs may also be seen due to the fact that people will have access to the software for themselves so there is no need to go through a middle man such as an estate agent.

### 3.II.b Legal and Ethical Matters

Throughout the world of the internet, many people have their personal data stored online. This could include from having a simple email account that is seemed to be secure to having a public profile on a social media site. Employers must actively check their employees’ actions on their computers to make sure no one is accessing files that could be deemed to harm the system. This directly could affect our own project with the Scrum Master and team members. Privacy has a lot of legal and ethical matters. For example, to what extent do people know that their accounts are being monitored in the workplace? Privacy in the computer industry can be easily seen as a slippery slope, slowing wearing down an individual’s right to privacy completely. [13]

One the main ethical issues would be in relation to the accuracy of our product. If the data used is sub standard or the code written for machine learning is sub-par then this would ultimately reflect in the exactness of the predicted prices. This is a major issue since it would effectively be misleading the user and could cost them a lot of money in the process. The result of this could lead to the user taking legal action for a product that effectively does not do what it states on the tin. These regulations can be enforced where someone who was “misled about the product or service you bought”[14] can be liable to compensation and taken up as far as trading standards.

In the world of software in the computing industry, people can access information to flow more freely than ever before. This action could be exploited and comes with a legal and ethical repercussion. The question can be asked to whether ownership can be established in the digital world. These days things can be easily copied and pasted online, which means that intellectual property is hard to control [15]. For example, when making a computer program and putting it out there in the internet, someone could easily copy your programming code and make a few changes and deem it as their own work. Legal notions such as copyright have struggled to keep up with the digital era. This matter could happen to our project at any time.

### 3.II.c Sustainability

Sustainability is a key aspect in maintaining an up to date and ultimately a successful long-term product. Sustainability can be thought of in multiple ways, one of them is in regard to the technical side to the product. Code concision is a must for being sustainable; real time updates is a big part of this and this can only be achieved with well written code. In our case, we decided to take a linear regression technique. By taking this approach, we are able to add a number of different features and independent variables on the go. This in effect will somewhat allow our machine learning to be constantly updated in accordance to real time market data. Without this, the product could become outdated fairly quickly due to the volatile nature of the housing sector. More specifically, housing desirability in terms of features is always changing too, so this must also be accounted for. One good example of this is the current Brexit ordeal. From February 2016 till august 2018, it can be observed that house prices have been on average gradually increase. However amongst the speculation and uncertainty, we can see that house prices have fallen in a short span of 4 months up until December 2018 [16]. These sorts of unforeseen events need to be somehow be considered when factoring in other elements if sustainability is to be maintained. A number of design factors have also got to be considered such as portability and robustness in order for it to be user friendly. In terms of the process itself, we took into account variable external factors such as minimising computer usage time to reduce our overall carbon footprint. This can be done via again, concisely written code so there is not a constant need to debug and re write code.

### 3.II.d Health and Safety Matters

In the world of Computer Science and related fields there are health and safety concerns that occur in the industry. One concern is computer stress that employee’s having to face in their day to day life when in the computing industry. Computer stress is caused by the direct result of using computers. In our product this will mainly link to the developers of the project who are having to spend hours on ends looking and working on computers, this could result in having the ability of not concentrating all the times, becoming angry towards the computer for crashing or losing data and the inability to work out how to solve a problem on the program [17]. These stresses can be relieved by doing simple acts to make sure they do not occur when working. For example, taking regular breaks when using the computer at work. Taking breaks is key when working for hours on end on designing a computer program, going outside for fresh air can refresh your mind and clear your mind from stress. Another way to relieve stress is being comfortable where you work. To make sure that you are sitting with your back straight and not slouching, to make sure your eyes are on level with the computer screen. These small acts can go along way when working with computers. [18]

Other health and safety matters include Repetitive Strain Injury (RSI), upper limb problems and backache from overuse or improper use of DSE (Display Screen Equipment). When designing a computer program, there are many examples of RSI. This can be caused by looking at a computer screen for a certain amount of time, sitting in a slouching position rather than having your back straight when working or using a keyboard at work and typing for a long amount of type can lead all lead RSI[19]. DSE (Display Screen Equipment) only affect workers who regularly use DSE as a significant part of their normal work. DSE are devices that have a graphical display screen, these include, display screen, laptops, touch screens and other similar devices. Workers who spend their normal work day using these devices could experience fatigue, eye strain, upper limb problems and backache from overuse or improper use of DSE [20]. The HSE [21] are currently the UK’s national watchdog for health and safety issues and this can be used as a guideline when creating a product like ours in order to maximise work efficiency whilst maintaining the health of the developers. In relation to our own product, we made sure this wasn’t a big issue since we used Jira to optimise workload and made sure everyone did as much they were able to do. For the end user, health and safety shouldn’t be a big problem for our product unless the user chooses to over use it and as a result stress over the devaluation or increase in the prices of homes.

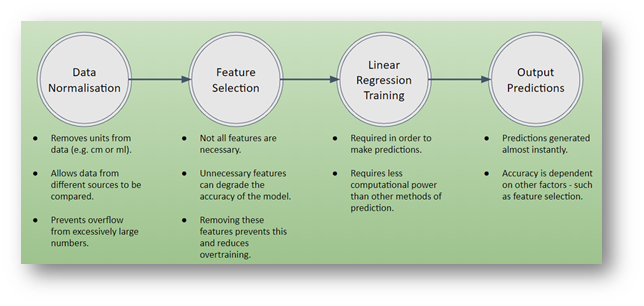
## 3.III The Team Product

## 3.III.a Product description

The House Price Prediction program uses data provided by Kaggle to generate predictions of house prices. The program normalises the data for analytical use, determines data that has the highest correlation to the price of a house, and then trains a linear regression model using this refined data. At the end of the prediction process, the program then writes these predictions to a file for submission to Kaggle.

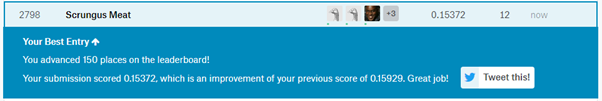
Kaggle provides two sets of data: the training data, and the test data. The training data contains all the same fields as the training data, in addition to the true price of each house. This makes the training data suitable for training the linear regression model. The test data is simply used to make predictions using the trained model and then associate each prediction with correct data – each entry in the test data contains an identifier, which must also be written to the submission file.

Generating predictions involves more than outlined above, however. There are several subsystems that are used to make the data processable by the linear regression model. Firstly, any categorical data (i.e. roof style) must be converted to their numerical equivalents. This is handled by the Python library, Pandas. Secondly, the data must be normalised – working on the data unnormalized could lead to large numbers (larger than the maximum computable by NumPy/Pandas) being generated. Data normalization is handled by a Standard Scaler (provided by Sklearn), which is first “fitted” to the data, before then being used to transform (normalize) the data. After the data is normalized, the correlation of each field to the house price is determined. The fields are then ranked in order of their correlation to house price, with the top 60-70 features being selected for training – removing features with low correlation reduces the chances of overtraining the model. Finally, the linear regression model is trained, the test data is processed (normalized after replacing categorical data), and the linear regression model is used to generate predictions.



*Figure 1: Shows an abstracted version of steps described above*

Our implementation was achieved through a trial-and-error process, whereby we continuously adapted our initial prototype until we were able to improve the score given to us by Kaggle. This process was repeated several times, with some iterations even involving major code changes. Despite being a trial-and-error process, it was by no means random – the changes we were making were educated. For each iteration, we explored why the score was not good enough and used information determined from that to focus our attention on specific areas of the code. For example, at the beginning of the project, we utilised a neural network; however, after attempting this approach and deciding that it was perhaps too complicated for us, we switched to a simpler, less complex linear regression model. Whilst this required a large portion of the code to be rewritten, it yielded far better results.

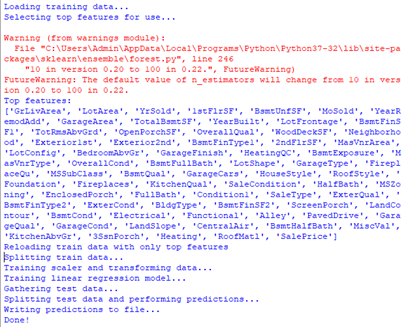


*Figure 2: Shows the rank given by Kaggle*

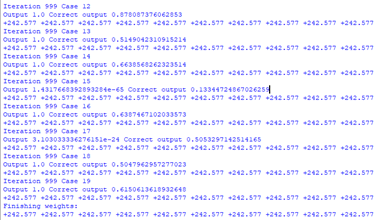
Using the linear regression model, we were able to obtain a rank of #2798 on Kaggle. This rank shows both our success and our room for improvement. With some improvements to the feature selection algorithm and the process of training the linear regression model, our rank of 2798 could be greatly reduced. Another area for improvement (and a limitation in itself) is the model used. For this project, we used a linear regression model; however, linear regression models are limited to linear data, and so changing this to a different mathematical model, or even a properly implemented neural network, also has the possibility of greatly improving the score.

## 3.III.b Product Demonstration

During the development of each iteration, we ran several tests to determine the accuracy of that version. One of these tests was making predictions on the training data, which allowed us to compare our prediction with the actual house price – something we were not able to do with the test data. Once we were satisfied with the results, or needed further clarification of those results, we submitted our predictions to Kaggle, after which we used the score given to us to determine whether we had improved the accuracy or not.

The latest version of the program describes the processes it is carrying out as it executes. This allowed us to check for any areas of the code that had a noticeably long processing time, in addition to providing debug information – outputting the top features, was especially useful.

**

Whilst the latest version of the code performs well and achieves the goal we set out, some earlier versions did not turn out so well.   
  
Figure 4 shows the output of an earlier version of the program, which behaves in a far more complex manner – using a neural network. This method was later replaced due to the high inaccuracy of the model – most likely due to our implementation.

# Chapter 4: Project Management (1044 words)

**4.I An introduction to Project Management**

A project is defined as a unique list of tasks that are connected in a sequence. Project management can be defined as the management of the required tasks.

There are five factors to be able to complete a report:

* Scope – it’s an agreement between the developers and the client. This is normally structured in a document.
* Quality - The quality of the product is how good or bad the product is
* Cost – How expensive is it going to be
* Time - time required to complete the project
* Resources – people and equipment

**4.II Project Management Report**

During the project, our team used the Scrum framework regarding our way of dividing up and organizing workloads. It was chosen due to being appropriate for the size of our team, for the use of the Jira platform during development and due to the flexibility of the framework itself.

The position of the Scrum master rotated on a meeting to meeting basis with the aim to have each member of the project hold the position at least once by the end. After a short discussion, the Scrum master split up the mentioned issues among the team and shortly outlined the tasks of each member in the following sprint. Additionally it was the Scrum master’s task to put these on Jira and to ensure that everyone understood what was expected of them during the sprint.

The sprints themselves were each two weeks long and were usually made up of tasks divided equally among all members of the team. Care was taken to not assign too little or too much to any single member during meetings. Additionally tasks were only given out after ensuring that the member in question understood what they were expected to do and how much work would have to be put into the task.

**4.II.a Summary of Teams Project management in Jira**

The online platform Jira was made specifically for the Scrum framework; as such it only made sense that it was to be used for the organization during the development of our project. It offers support for the main features of Scrum, such as the ability to assign tasks, add them to sprints and then start sprints for specific lengths of time. Additionally it allows for close monitoring of progress and tracking of overall participation if used correctly.

During the time of our project, the team used Jira extensively. Each meeting a new Scrum master was given the responsibility of assigning tasks on the platform and members were expected to, while taking care of whatever task had been assigned to them, track their progress to the best of their abilities. Ideally, members would log onto Jira whenever any work was done on the project and denote as much for the convenience and information of all other members.

One issue members sometimes found themselves running into was incorrectly tracking their time spent. They would work on their tasks for some time, however only log onto Jira to denote when they finished, rather than correctly logging each time the task was actively worked on. This caused the various graphs to not accurately reflect the actual activity of the group, but instead made it seem like tasks were only worked on and completed closer to the end of a sprint (Which was not necessarily the case). Essentially this meant that to some, Jira became more of an afterthought when compared to the actual completion of tasks. While this is understandable, it nonetheless led to issues with correctly approximating the member progress.

Overall the utilization of Jira did help the team organization, as it gave a commonly accessible space to everyone to check on what they should be working on at any given moment. However our use of the platform could have been improved by sticking closer to its intended use of consistent logging of work. The issues stemming from lack of proper documentation should have been addressed in a stronger way. Nonetheless, apart from the lacking documentation during sprints, the tasks themselves were usually completed on time and each member fulfilled their expected roles whether that be Scrum master or simply working on a task during a given sprint. To conclude, Jira was a powerful platform used well by our team, however not to its full capacity.

**4.II.b An evaluation of the Project management**

As outlined above, the management of the project was handled by the Scrum master at any given time (With input from every member as to play to each other’s strengths and weaknesses). The position of the scrum master changed week by week as to ensure that everyone held a position of greater responsibility at least once.

Each member handled their position as the Scrum master well. The help of the Jira platform coupled with the fact that the group held thorough discussions during meetings when it came to assigning tasks ensured that the position of the Scrum master was neither stressful nor difficult. As such, the transition of the position from one member to the next was always smooth and unproblematic. Similarly, the job done by the scrum master always produced a sprint that gave full understanding to members in terms of what was expected of them. Vital in this was the communication and positive work environment within the group, which ensured that everyone felt comfortable voicing even the smallest issue before it could grow into an actual problem. While communication was reduced during sprints, this was often since a well-executed meeting gave everyone an understanding of what to do, rather than the members themselves feeling unsure on their tasks but being uncomfortable voicing them.

The performance of each member when it came to both leadership and non-leadership roles allowed for a smooth development. The only issue that could be raised would be the lack of experience in terms of development time, which can be often hard to predict as unexpected issues can arise at any moment and hold up work for an unforeseeable amount of time. Therefore it would be difficult to fault leadership in an area we cannot expect them to have experience in beforehand.

Overall, the leadership within the project was smooth and issueless with a high amount of communication and a decent level of understanding.

# Chapter 5: Conclusions (334 words)

Overall, we did work well as a team. We worked hard from the start and was able to produce a solid program that worked. Despite the fact that we all had different levels of knowledge when it came to programming and machine learning, we able to learn from each other and become stronger coders in our own right. We had also met the requirement of the project. Our team organisation was also ok, we were able to quickly choose our scrum master and assign individual tasks effectively. Due to this, we were able to finish the project on time, however it wasn’t without their issues.

Setting up meetings outside of the lab times proved difficult as everyone wasn’t available at the same time and some of us did not live on campus, thus making it difficult hold a meeting at a time which was suitable for everyone. This prevented us from having more meetings which possibly stopped us from creating a better product. We also had problems with signing off tasks in Jira. Most times, we would complete a task and not mark it down on Jira, which means that others cannot tell if people have completed tasks or not. The second problem was not holding enough meetings outside of our lab. Although we were on campus at different times, we still could have held more meetings outside of our labs (especially during the beginning). It really showed towards the end when we got more done due to these meetings. Our final problem was during the second sprint. Victoria had overestimated the time it would take people to complete their tasks which was one of the reasons that the due date of the sprint was extended. Another reason was that one person had fallen ill and was unable to work on their part, which was vital to this project.

If we were able to have further development, we would possibly use multiple linear regression as it is more accurate due to using multiple predictors.

References

[1] Wikipedia, Definition of Teamwork, <https://en.wikipedia.org/wiki/Teamwork>

[2] Book 2, Chapter 7, Patrick Lencioni: Five Dysfunctions of a Team, <https://moodle.essex.ac.uk/mod/book/view.php?id=434961&chapterid=10129>

[3]Book 2, Chapter 5.1 Tuckman, <https://moodle.essex.ac.uk/mod/book/view.php?id=434961&chapterid=10126>

[4]Book 2, Chapter 6 Behavioural team roles - Belbin, <https://moodle.essex.ac.uk/mod/book/view.php?id=434961&chapterid=10128>

[5]MindTools.com , Belbin’s Team Roles - Leadership Training, <https://www.mindtools.com/pages/article/newLDR_83.htm>

[6]"Multicollinearity - Statistics Solutions", Statistics Solutions. [Online]. Available: <https://www.statisticssolutions.com/multicollinearity/> . [Accessed: 24- Mar- 2019].

|  |
| --- |
| [7] K. a. Dawson, Software Project Management, Group D Publications Ltd, 1999. |

|  |
| --- |
| [8] I. K. Portal, "Software Development Methodologies," IT Knowledge,  [Online]. Available: <http://www.itinfo.am/eng/software-development-methodologies/>.[Accessed 20 March 2019]. |

|  |
| --- |
| [9] S. E. Team, "Top 4 software development methodologies," Team Synonpsys, 28 March 2017. [Online].Available:<https://www.synopsys.com/blogs/software-security/top-4-software-development-methodologies/>. [Accessed 21 March 2019]. |

[10] U. o. Essex, "Team Challenge Project Book 3 Product Development," [Online]. Available: <https://moodle.essex.ac.uk/mod/book/view.php?id=464360>. [Accessed 23 March 2019].

|  |
| --- |
| [11]A. Sergeev, "Team Roles in Waterfall Methodology," 12 February 2016. [Online]. Available: <https://hygger.io/blog/team-roles-in-waterfall-methodology/>. [Accessed 21 March 2019]. |

[12] S. W. Ambler, “Roles on Agile Teams: From Small to Large Teams,” Ambysoft, 2012. [Online]. Available:<http://www.ambysoft.com/essays/agileRoles.html>.

|  |
| --- |
| [13]C. Hope, "Privacy," 17 October 2017. [Online]. Available: <https://www.computerhope.com/jargon/p/privacy.htm>. [Accessed 22 March 2019]. |

[14] C. Advice, "If you were misled or pressured into buying something you didn’t want," Citizens Advice, 2019. [Online]. Available: <https://www.citizensadvice.org.uk/consumer/changed-your-mind/if-you-were-misled-or-pressured-into-buying-something-you-didnt-want/>. [Accessed 2019 March 24].

|  |
| --- |
| [15] Jisc, "Intellectual property rights in a digital world," 17 August 2015. [Online]. Available: <https://www.jisc.ac.uk/guides/intellectual-property-rights-in-a-digital-world>. [Accessed 22 March 2019]. |

[16] [2] E. Clark, "What will Brexit mean for house prices?," Which.co.uk, March 2019. [Online]. Available: <https://www.which.co.uk/news/2019/03/what-will-brexit-mean-for-house-prices/>. [Accessed 24 March 2019].

|  |
| --- |
| [17]S. R. Choices, "Computer Stress," Stress Relief Choices, [Online]. Available: <https://www.stress-relief-choices.com/computer-stress.html>. [Accessed 22 March 2019]. |

|  |
| --- |
| [18] S. R. Choices, "7 Great Tips for Computer Stress Relief," Stress Relief Choices,[Online]. Available: <https://www.stress-relief-choices.com/computer-stress-relief.html>. [Accessed 22 March 2019]. |

|  |
| --- |
| [19]H. a. S. Executive, "Managing upper limb disorders in the workplace," Health and Safety Executive, August 2013. [Online]. Available: <http://www.hse.gov.uk/pubns/indg171.pdf>. [Accessed 23 March 2019]. |

|  |
| --- |
| [20] H. a. S. Executive, "Working with display screen equipment (DSE),"Health and Safety Executive, May 2013. [Online]. Available: <http://www.hse.gov.uk/pubns/indg36.pdf>. [Accessed 23 March 2019]. |

[21] HSE, "HSE," HSE, 2019. [Online]. Available: <http://www.hse.gov.uk/index.htm>.[Accessed 25 March 2019].