TECHNISCHE UNIVERSITÄT CHEMNITZ

Faculty of Computer Science

Prof. Dr.-Ing. M. Gaedke

Planspiel Project Report

X-Review

Team name: WebFabrique

Chemnitz, 28. März 2018

Team members	**Matriculation no.**
Monty Bhattacharyya	452625
Bojan Kovachki	455598
A K M Tufazzul	456649
Orçun Oruç	321448

Abstract

The goal of our project is to offer manufacturers, retailers, wholesalers and regular online shoppers the ability to effortlessly access customer review analysis, historical price analysis, price drop prediction and price comparison of a product that is offered on a certain e-commerce website (such as Amazon and eBay).

We provide our customers with a product's price deviation and offer a sentimental analysis of the product. This is done by using cutting-edge techniques such as LSTM, ARIMA and Natural Language Processing.

The product's feature extraction that our software is offering may direct future research on characteristics feature extraction with unsupervised learning algorithms.

We hope that our work will provide valuable information to prospective researchers in the areas of Big Data, e-commerce and Machine Learning.

Table of Contents

List of Figures

Log book

Month: October, November, December, January, February, and March
(16/10/16 – 28/03/16)

No	WEEK	DURATION	RESPONSIBLE PERSON	ACTIVITY
1	Week 01	16th Oct - 22nd Oct 2017	Team	1. Discussion about team goals after team formation.
			Team	2. Discussion about availability of team members throughout semester.
			Monty	3. Setup Bitrix24 Project management tool for planspiel and create timeline for major tasks throughout semester.
			Bojan, A K M, Orcun	4. Setup individual bitrix accounts and accept invitation for planspiel project
			Monty	5. Create all members as moderators for the project.
			Team	6. Discussing and sharing ideas about the team name and an adequate logo for it.
			Team	7. Voting for the team name/logo.
			Team	8. Creating the story and the message that the team name/logo broadcasts.
			Orcun	9. Developing the team mission and vision.
			Team	10. Writing them down, making them shorter and clearer and refining them.
			Team	11. Defining a website template that fits our team name and logo and that displays our technical abilities.
			Team	12. Distributing work to team members for website creation.
			Monty, A K M	13. Work on Creating the Website using template.
			Bojan	14. Creating the Logo of the company
			Orcun	15. will work on defining team mission, vision and statement.
2	Week 02	23rd Oct - 29th Oct 2017	Team	1. Finalization of the logo.
			Team	2. Voting for confirmation of Logo
			Team	3. Brainstorming on possible website layouts.
			Team	4. Voting for Template for company webpage.
			Team	5. Finalization of template.

3	Week 03	30th Oct - 05 Nov 2017	Team	1. Discussion about the website's content and the position of the logo, vision, mission and goal in the website.
			A K M	2. Refine CSS of webpage for a better look and feel.
			Orcun	3. Installing and configuring Bitbucket for source controlling, give read/write access to all members.
			Team	4. Create team member account on Bitbucket.
4	Week 04	6thNov - 12th Nov 2017	Team	1. Finding out the challenges using the uk gov template.
			Team	2. Brain storming on a problem statement for LFA.
			Monty, A K M	3. Create WebFabrique company website using the uk gov template.
			Team	4.Voting and Finalizing the problem statement.
			Bojan	5. Create LFA, problem tree and objective tree.
			Orcun	6. Final deployment to Heroku of our website (designed with and according to the UK Gov's website tool).
			Team	7. Making conclusions about out implementation of LFA.
			Team	8. Prepare for a short presentation about webfabrique.
			Team	9. Checking if our Heroku deployment works fine.
			Team	10. Short preparation for our presentation about it and the LFA problem trees.
			Team	11. Team: Watching videos (including the one suggested by Prof. Dr. Gaedke) about pitching.
			Team	12. Sharing and exchanging personal pitching experiences.
			Team	13. Search for ideas how to refine pitches.
			Team	14. Preparing our first pitch and doing some actual pitching to random people in a local club.
			Team	15. practice the pitching on a group level along with gestures and body language
			Team	16. Finessing our initial pitch and delivering the improved pitch to fellow students at the student dormitory.
5	Week 05	13thNov - 19th Nov 2017	Monty	1.Knowledge transfer with the Bitrix management tool and gives hints and tips on how to use it to the rest of the team.
			Team	2. Setup bitrix account

6	Week 06	20th Nov - 26th Nov	Monty, A K M	1. Adding the excel log book file to the website and making changes to the website source.
7	Week 07	27th Nov-3rd Dec 2017	Team	1. Discussing the possible approaches and solutions to the problem we are going to get (academical or industrial).
			Team	2. Finalize the technologies we might use for the development process.
			Team	3. Setup repositories for final project.
			Team	4. Setup environment for development using the finalized technologies.
			Team	5. Talked about the ideas provided by prof.
8	Week 08	4th Dec - 10th Dec 2017	Team	1. Preparation of the question about the idea offered by professor
			Team	2. Generating ideas about the topics got from prof.
9	Week 9	11th Dec-17st Dec 2017	Team	1. Choose the topic of Big data and Ecommerce
			Team	2. Email to Bahareh on 20th we choose Big data and Ecommerce
	CHRISTMAS VACATION	18th Dec-31st Dec 2017		
10	Week 10	01th Jan -07th Jan 2018	Team	1. Research on some tools to make a prototype
				2. Understanding of Big Data Topic from Professor Gaedke.
				3. Gather understanding of basic machine learning
				4. Study Data gathered from IBM Big Data Analytics tool
				5. Research on Magento Open source framework
				7. Read on Big Data Analytics and its importance in industry
11	Week 11	08th Jan -14th Jan 2018	Team	1. Brain storming sessions to analyse idea provided by professor and find new key requirements.
				2. Gather ideas to reach out to customers
				3. Survey idea to reach companies and get potential customers
				4. Reach out to customers for Phase1 survey
			A K M, Bojan	1. Create prototype using Balsamiq Software to be sent to customers

			Monty	1. Formulate questionnaire for customers for big data analytics usage in companies
			Orcun	2. Create detailed customer journey map, as a click-able pdf to be shared with customers
12	Week 12	15th Jan -21th Jan 2018	Team	1. Brain storming sessions to analyse idea provided by professor and find new key requirements.
				2. Find companies email id to send survey questionnaire
				3. Brainstorming with ideas and realized a new and unique idea for planspiel 18th January.
			Orcun	1.Feature extraction for sentiment analysis on customer reviews
			Bojan	1. Data crawling from Amazon, eBay for products
			Team	Market and send survey to customers and gather feedback
			A K M	Designed the initial product architecture
13	Week 13	22th Jan -28th Jan 2018	Team	1. Market and talk to customers about phase 1 survey, and try to gather feedback and data
			A K M	1. Designed the outlook of the dashboard- price comparison, feature extraction using bootstrap and html5
			Bojan	1. General Data Protection Regulation (GDPR) study and rules compilation
			Orcun	1. Create Business canvas model for product
			Monty	1. Survey preparation and result summarization
				2. Create Google View Analytics View
14	Week 14	29th Jan -04th Feb 2018	Team	1. Market and talk to customers about phase 1 survey, and try to gather feedback and data
				2. Extract results from Phase 1 survey from companies to define Idea
			Monty, Bojan	1. Discuss and formulate questions for Phase 2 survey
				2. Discuss Phase 2 survey marketing strategy by using social media
			Orcun	1. Created product video to be used in marketing Xreview to customers
			Monty	1. Create Phase 2 survey using google forms
				2. Created animated GIF's from dashboard to be used in product video
				3. Resolve resolution issue in product video

			Bojan	1. Automate WebCrawler to automatically extract information for specific products
			A K M	1. Create initial business plan to be shared with Prof Gaedke
15	Week 15	05th Feb -11th Feb 2018	Team	1. Prepare Meeting with Prof Gaedke
			Bojan, Orcun, AKM	1. Create swot analysis and combine outcomes to formulate strategy for team
			Bojan	1. Create crawler to extract product info from Amazon
			Orcun	1. Create code for analysing sentiments using NLP gathered from data extracted from crawler
			Monty	1. Setup NPM & gulp code for angular for initial dashboard setup
				2. Setup dashboard layout and design for presenting to professor
				3. Compare different tools like Clara analytics to find our X-factor or differentiator
			A K M, Bojan	1. Created business plan and costing structure with forecast for next 5 years
16	Week 16	12th Feb -18th Feb 2018	Team	1. Marketing customer survey to customers using social media
				2. Plan aggressive strategy usage for marketing, from the results of Swot Analysis
			Bojan	1. Optimize Dexi web crawler 2. Start creating own web crawler with Python
			Orcun	1. Refinement of Text Extraction from customer reviews
			Monty	1. Xreview dashboard application user page design with profile info and customer social media information
				2. Xreview dashboard application Notification tab design and layout
				3. Dashboard Login and password page
			A K M	1.Price data analysis with python
				2.Fixed some gulp issue for the frontend
17	Week 17	19th Feb -25th Feb 2018	Team	1. Marketing customer survey to customers using social media
			Bojan	1. Build and train machine learning models with Mateverse. 2. Share Machine Learning knowledge with Monty and Tapu and suggest a book for them to read:

				https://www-01.ibm.com/common/ssi/cgi-bin/ssi-alias?htmlfid=IMM14209USEN 3. Finish development of the web crawler with Python (to be used in the future)
			Orcun	LSTM (Long-Short Term Memory) algorithm has been placed with ARIMA
			Monty	1. XReview dashboard Application Product Directory page design and data extraction from Orcun webservices
				2. Xreview dashboard Application Add Products to application for customers
				3. Xreview dashboard Application Notification and Mail layout design
			A K M	1. Compare different tools for price prediction and evaluate our product
18	Week 18	26th Feb -04th March 2018	Team	1. Prepare for Meeting with VSR team 2. Marketing of survey to potential customers 3. Idea generation for final pitch and plan to discuss with VSR group 4. Question formulation for final day
			Monty, Orcun	1. Plan to integrate price prediction to dashboard 2. Discuss, update and refine Magento idea for final pitch 3. Performance VM setup and access.
			Monty, Bojan	1. Refine Product video and add customer reactions
			Bojan	1. Create business model and costing structure 2. Finish the Google Machine Learning Crash Course (https://developers.google.com/machine-learning/crash-course/)
19	Week 19	06th March - 11 March 2018	Bojan, AKM	1. Work on business model by extracting data from phase 2 survey
				2. Refine model after discussing with team
			Monty, Orcun	1. Plan integration of services from dashboard to backend machine learning algorithms
			Orcun	Recurrent Neural Network testing with TensorFlow(Implementation)
				Test implementation and webservices

			Monty	1. Integrated Products from amazon data from orcun webservices
				2. Feature extraction dashboard creation from machine learning results
				3. Integrated BrowserSync with Dashboard to have Realtime data sync across browsers
20	Week 20	12th March - 14 March	Team	1. Select template for presentation
				2. Use Prezi tool to create individual pages for presentation
				3. Prepare pitch for final presentation on 14th
				4. Brainstorm new ideas for final presentation
				5. Discuss presentation for individual roles and mission of company
			Bojan	1. Create flyers for final pitch
			Monty, Bojan	1. Create final day Quiz survey and amazon gifts
				2. Flyers assemble
			Orcun	1. Print Team T-shirts
			A K M	Analysed content for final pitch
21	Week 21	15th March - 28 March 2018	Team	1. Report Creation

Vision/Mission Statement and Goals

Vision Statement

To be a globally respected web-based software company with its top-notch employees that provides state-of-the-art solutions by leveraging the latest web engineering technology.

Goals and Objectives

Provide a quality service and assistance to our customers at least equal to the highest standards in the industry and ensure that our customers are well informed about the use of our products.

Mission

Provide innovative web solutions to our customers and make them easily integrate in today's business world.

Company Summary

We started our startup as an academic project. Within two years, we would like to see our company in the top 1000 companies in Germany.

Management

Monty Bhattacharyya, CEO

Mr. Bhattacharyya is responsible to implement the strategic goals and objectives of the organization. To give direction and leadership toward the achievement of the organization's mission, strategy and its annual goals and objectives.

A K M TUFAZZUL, COO

Mr. Tufazzul is responsible for providing strategic leadership and direction to the company's daily operation. Ensure appropriate systems, processes and performance management arrangements are in place to deliver consistent high-quality level of service provision and activity report and monitor achievement.

Bojan Kovachki, CFO

Mr. Kovachki is responsible for analyse, forecast, and project financial trends in the company. Create and monitor an annual company budget to serve as a roadmap for company growth. Development of corporate policies related to best practices and regulatory compliance.

Orçun Oruç, CTO

Mr. Oruç is responsible for ensuring the speed, accuracy, availability, throughput and reliability of IT. Involved in the creation of new technology and its application to company products and services.

Business Canvas Model

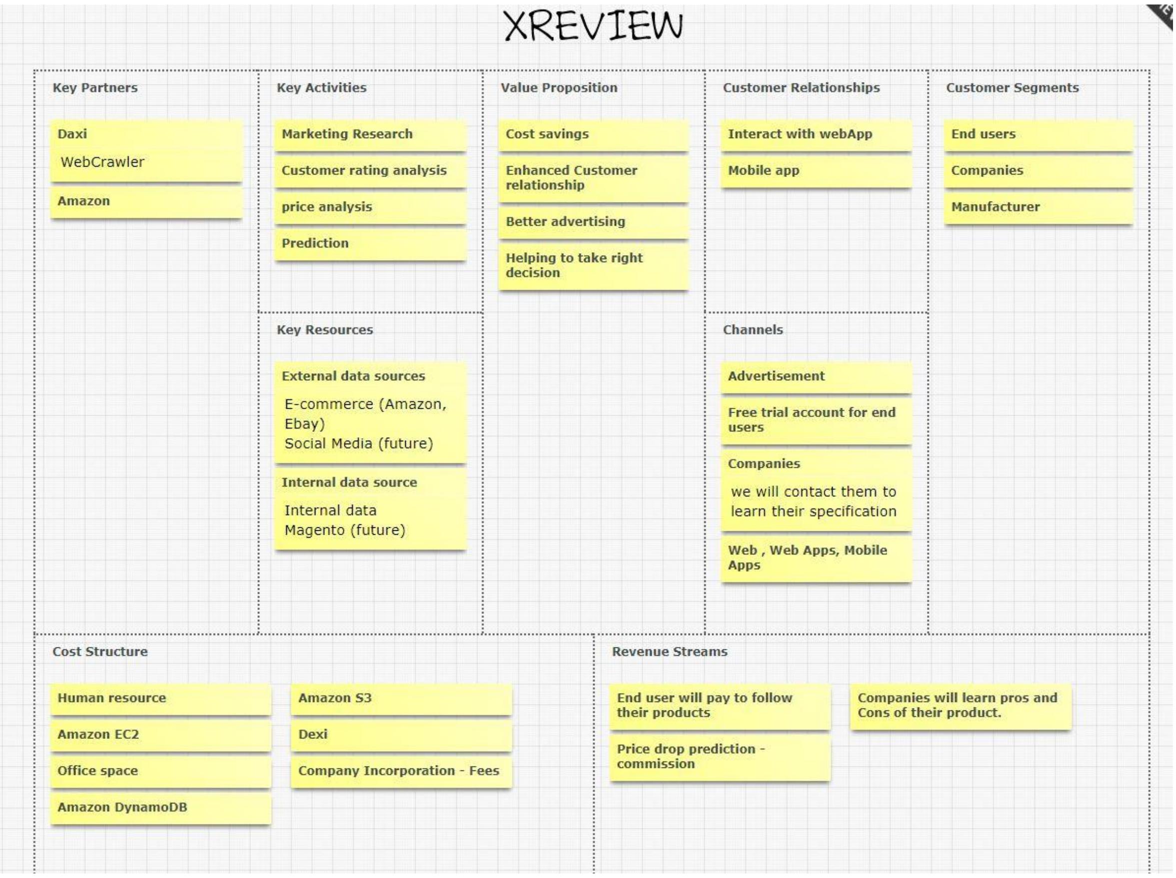

Figure 1: Business Canvas Model

Products and Services

Customers can follow their desired products and they will get updated about these products. In the dashboard, the customer gets all the important information for a certain product. We are displaying the current price of the chosen product from different market places (price comparison among different e–commerce market places). Our customers also get a feature extraction for the product. For example, a customer starts following a Canon 750D DSLR camera. In the feature extraction part, he/she will see different charts in which it is clearly visible which are the positive and the negative aspects of this product. We are extracting these features from customer comments left on various e-commerce websites.

The feature extraction offers a thorough and legitimate review of the product. In addition, we offer another great option in our product which is price drop prediction. We are analysing the history of the product's price and by using machine learning algorithms we are predicting when the price of a certain product is going to drop. Our customers get a notification when there is a price drop for the product they are following.

Marketing Methods

webFabrique will market X-Review through the following methods and sales channels:

- Posters
- TV ads
- Social Media
- Email Marketing
- Facebook Advertising, Google My Business, Google AdWords, Content Marketing, Organic Social Media, Coupon Deal Sites, Speak at Events

Ownership Structure

WebFabrique is organized as a Limited Liability Company (LLC) in the state of Saxony in Germany since October 2017. The purpose of this business plan is to raise 200 000 euros to finance the founding of our new business.

Monty Bhattacharyya, CEO

Ownership Interest: 25% common stock.

A K M Tufazzul, COO

Ownership Interest: 25% common stock.

Bojan Kovachki,CFO

Ownership Interest: 25% common stock.

Orcun Oruc, CTO

Ownership Interest: 25% common stock.

WebFabrique is actively looking for investors; the aforementioned chairmen of the company will each release 5% of their respective shares and give it to the investor who offers to help create the startup with an investment of 200 000 euros.

Internal Analysis (SWOT Analysis)

Strengths:

- Unique product
- Great team spirit
- Highly skilled staff

Weaknesses:

- Small team
- Low brand recognition
- Software not useful for cheap items

Opportunities:

- Emerging interest in the market for such a product
- Enlarging the scope of the product (e.g. airline tickets)
- Diverse sources of revenue

Threats:

- New competitors
- Regulatory hurdles
- E-commerce companies (Amazon, eBay etc.) manipulating prices order to harm us

Swot Conclusion

- **Strategy:** Aggressive

What we should do:

- Make use of internal strengths and focus on utilizing opportunities to gain an advantage
- Invest heavily in innovation by creating new solutions
- Cover any moves made by competitors
- Explore emerging markets which can be a complement to our current ones
- Raise the stakes; even reduce our profit margin if necessary

Market Assessment

Customer Analysis

1. Who are our customers?

- Our customers are product manufacturers, retailers, wholesalers and individuals who do online shopping.

2. What do we sell to each of the customers?

- Product manufacturers, retailers and wholesalers mainly benefit from the customer review analysis. The online shoppers' main benefit derives from the historical price analysis and the price drop prediction.

3. How does your product/service solve a key customer problem?

- Our product solves two main customer problems:
  - Feature extraction from products (positive/negative reviews)
  - Price analysis (the customers want to know what the best time is to buy a certain product and from which seller)

4. How difficult is it to retain a customer?

- Retaining a customer for us is not an issue, since our app has a subscription model which includes a free account and in addition the customer feedback is very positive (over 90% of our customers like to use the app and would recommend it to others).

Market Analysis

We have analyzed data offered by the Statista Market Forecast. According to their numbers, the Average Revenue per User (ARPU) for online shopping in Germany in 2018 equals to the amount of 1093.90 euros. This number is the sum around which our business plan revolves.

We do not want to force the online shoppers to spend more, but on the contrary, we want to enable them to save some money from the sum they are already spending. Our initial goal is to earn at least 1% of the ARPU, meaning that we are trying to earn some money from the amount that the users are already spending (instead of making them spend additional money for our product).

By reducing the basic ARPU which tends to grow over the years, we are hoping that the gap that is being created is going to end up as a profit for our company. The philosophy is simple: People save money thanks to our app (via price drop prediction), so they have some extra money to spend on more products (by spending money on more products, we are earning more, because we charge our customers 5% of the discount for each product when our price drop prediction is correct).

Competitors

webFabrique is offering a complete package to our customers. There are other products available on the market which have certain features. Our product X-Review is the full package – it contains all the features which are offered by our competitors and a few extra ones. The following is a table showing a comparison between X-Review and other products in the market.

Feature / Company Name	WebFabrique	pricefrontier	Pricify	Airhint
Customer review Analysis	Yes	No	NO	No
Historical Price Analysis	Yes	Yes	No	No
Price Comparison	Yes	No	No	No
Price Drop Prediction	Yes	No	No	Yes
Price Drop alert	Yes	No	Yes	No

Figure 2: Competitors Comparison

Financial Plan

Pricing

Our pricing model has three subscription options:

- Free
- Standard
- Business

The following is a visual representation of the subscription categories:

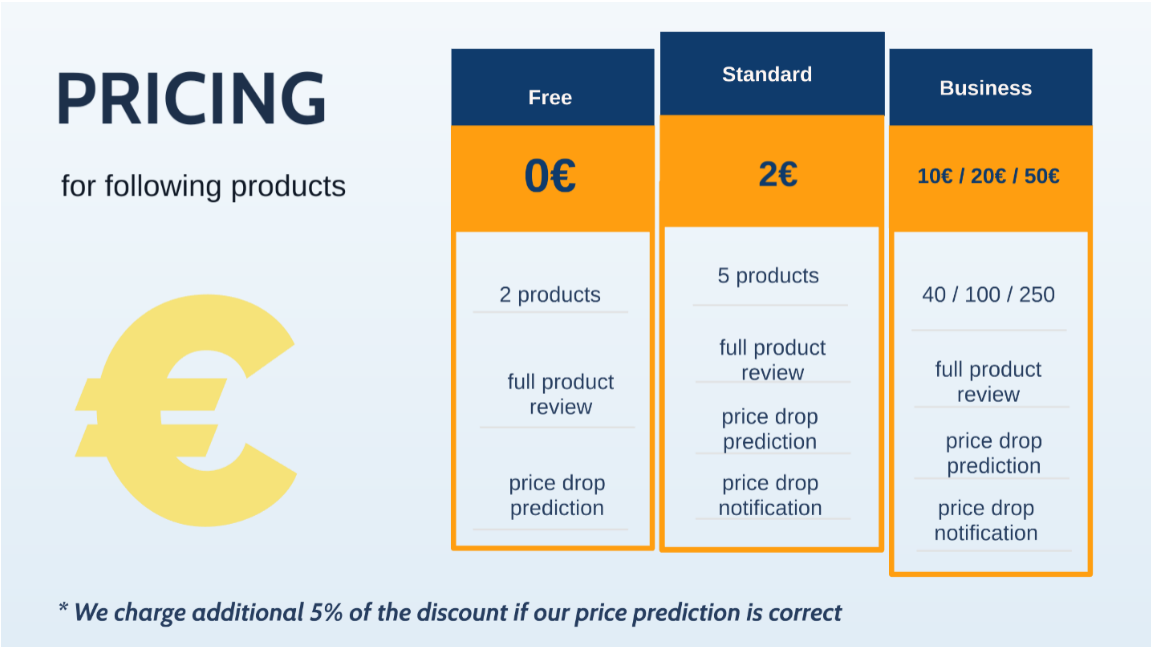

Figure 3: webFabrique pricing (subscription) model

As can be seen in the previous figure, we are charging additional 5% of the discount of each product if the price prediction is correct. The idea is to motivate the users to try/use our app more, since they can only save money by using it (the users are not charged anything if they opt for the free subscription model and if they do not benefit from our price prediction algorithm).

First Year – Income and Expenditure

This paragraph is dedicated to the financial aspect of our business operations during the first year of our existence. In the next two figures, we are going to have a look at a summary of the basic incomes and expenses that are going to appear in the initial twelve months of our operation.

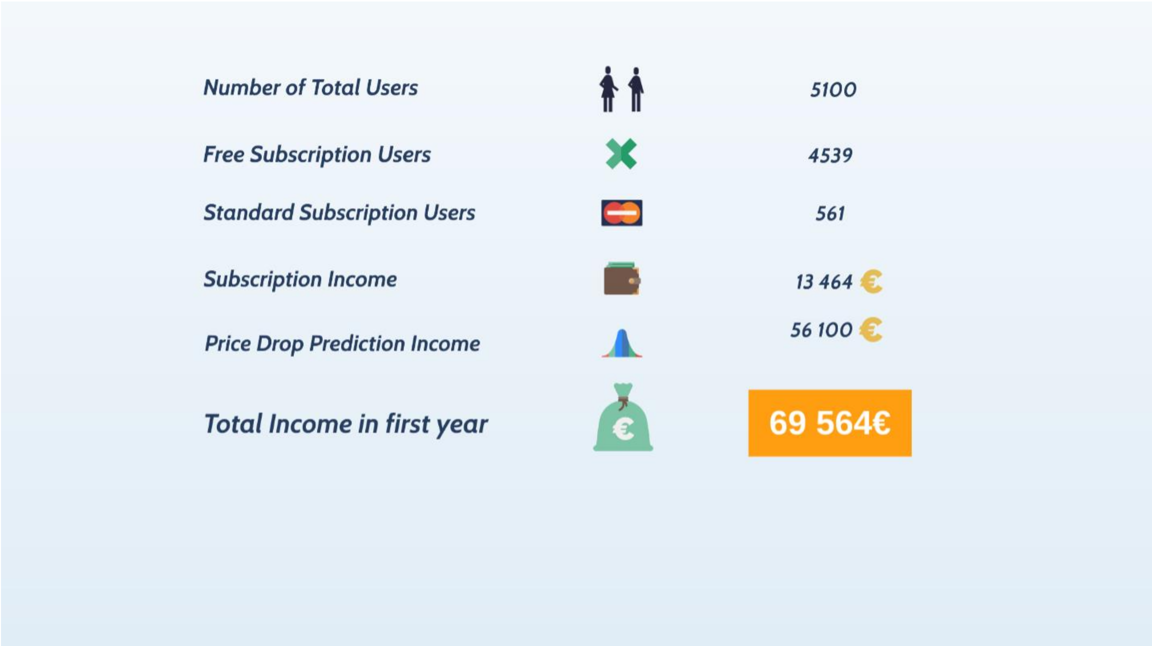

Figure 4: webFabrique first year income

Figure 4: webFabrique first year expenditure

WebFabrique Surveys and Article

webFabrique values the opinion of its customers. Therefore, we have created two surveys whose goal was to shape and improve our initial ideas about our project.

The first survey (survey version 1) was tackling the issue of Big Data. The target of the survey were professionals mainly working in the IT world, specifically in software development companies. We went one step further and put some additional effort to contact as many people as we could who already had some experience in Big Data. Thanks to this survey, we were able to come up with and improve our idea about our product (X-Review).

The second survey (survey version 2) contained concrete questions about X-Review. This survey helped us perfect our business plan. It was generally targeting online shoppers, although there were some people who answered that they did not like to shop online.

Surveys' Strategy

Driven by the goal of reaching out to as many people as we possibly could, we developed a strategy that contained specific steps which needed to be taken. This strategy was revolving around the idea that the easiest way to contact people nowadays is through social media.

The first step was to share the duties among the four members of our team. Each one of us received a task he needed to execute. We all had to do a small research of social media groups that contained a significant number of members that were IT professionals (for the first survey) and of members who are online shopping aficionados (for the second survey).

The second step was to post information, questions and the survey itself in these groups.

The third and final step was to keep following the development of the discussion in these groups and keep engaging with the people that are answering the survey.

The fourth and additional step which was added later (out of necessity) was to analyse all the comments we are receiving in the survey and shape our business plan/product by using them.

Thanks to this strategy, we were able to reach out to 535 individuals who gave their opinion on various questions.

58 out of 535 individuals were surveyed in the first phase, and the remaining 477 were surveyed in the second phase. These numbers served as a solid base for our future planning in terms of market growth and finances.

Survey Version 1 (58 of 535 replies)

As mentioned previously, the first survey was the Big Data survey. This survey gave birth to our product – X-Review.

We are now going to have a look at two of the key questions that helped us come up with the idea for our product.

1. What do you think is the biggest opportunity for using Big Data in your company? Please give an example.

The most common answers given to this question were:

- Customer analysis
- Estimating market trends
- Focus on target groups to increase revenue
- Cost reduction, scrap reduction, improving product quality, improving decision making

2. Do you have any Big Data initiatives in progress or in planning?

Here is the chart visualization of the received answers:

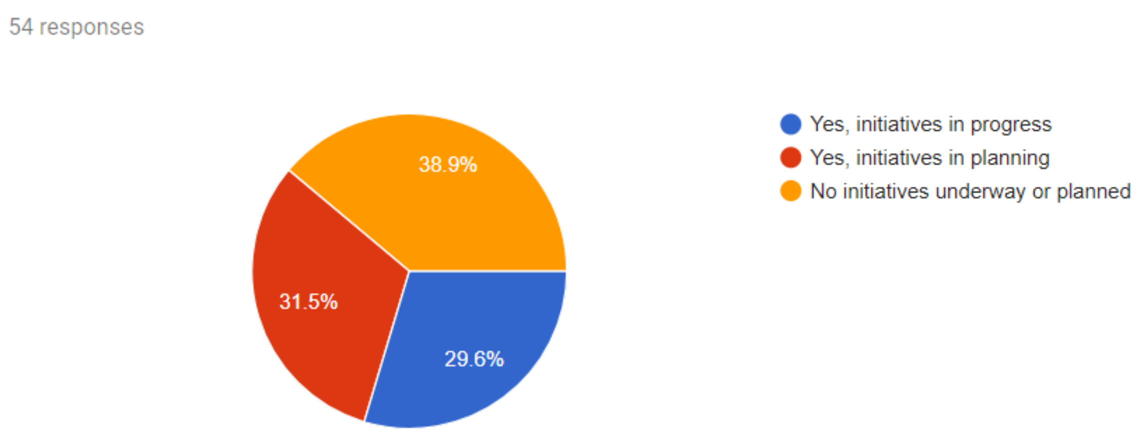

Figure 5: Big Data Opportunity Summary

Outcome
Carefully analysing the given answers, we realised that although the use of Big Data in the companies is rising, it still has not reached the sufficient level. Therefore, we decided to lead the way by creating an e-commerce product that benefits using Big Data and machine learning (Long short-term memory and Natural Language Processing).

Survey Version 2 (477 of 535 replies)
The second survey was the one that shaped our business plan - by asking questions about the pricing model, we managed to get valuable information and feedback regarding this subject.

Furthermore, we also asked questions regarding e-commerce and online shopping habits.

Here are some of the key questions (and their adequate answers) from this survey:

1. Would you wait for a product to be discounted before you buy it?

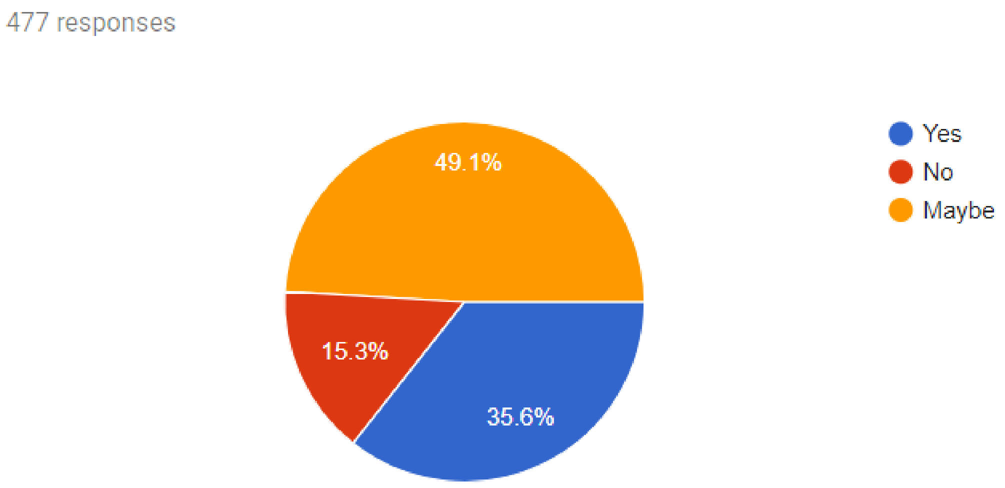

Figure 6: Product Discount Summary

2. Do reviews and rating (positive/negative) of a product have influence on your purchase decision?

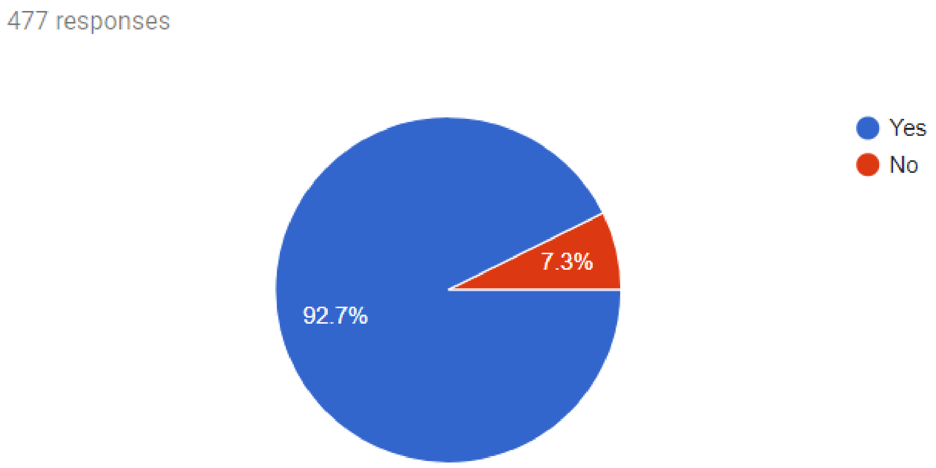

Figure 7: Review and Rating Influence Summary

3. Would you like to use a software that predicts the product's price (lowest price) along with a thorough product review analysis?

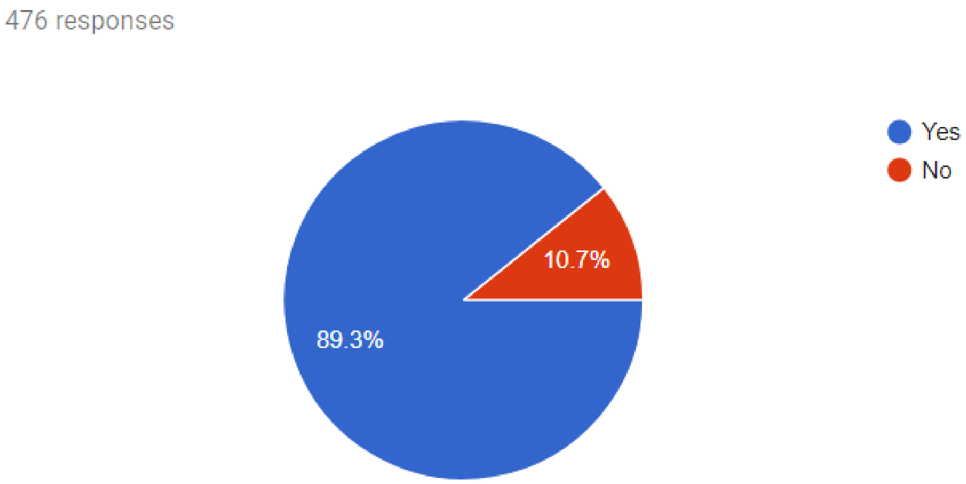

Figure 8: Customer Interest Summary

4. Which business model do you prefer from the following:

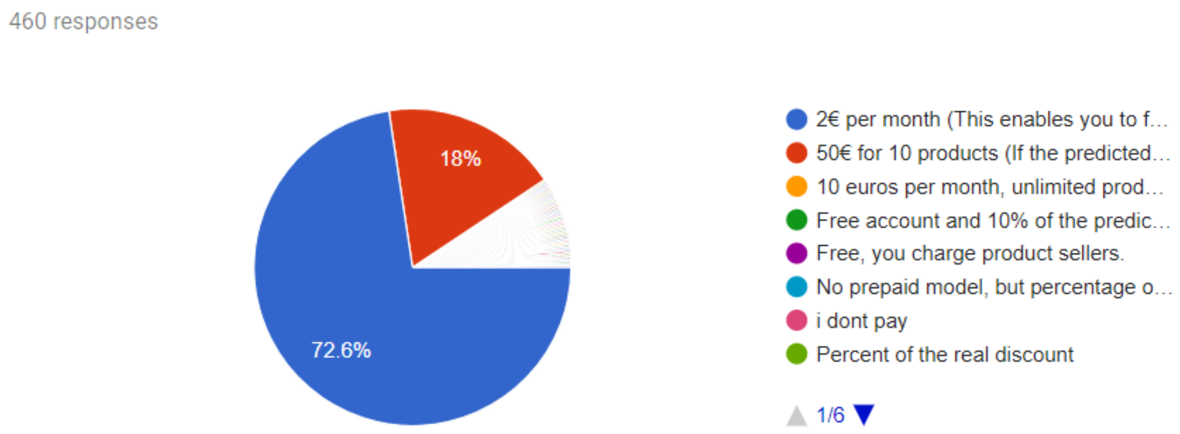

Figure 9: Business Model Summary

Outcome

The answers given to the questions in this survey are representing a cornerstone for our business plan. An additional question which was asking the surveys to leave a comment about the product or the business plan proved crucial. A significant number of people left profound and interesting comments, offering their help and suggestions on how we can improve our product and/or our business plan and marketing strategy.

Smartportal.MK Article

We consider our survey strategy a success. The requirement set by our potential investor (professor Gaedke) was to reach out to at least 500 individuals.

We managed not only to accomplish this goal, but also to overachieve (535 replies in total).

A significant role in this achievement was played by an online article that was published on a Macedonian website dedicated to smart technology. As a direct result of this article, we had collected around 150 replies. The article contained links to our product video and to our survey (version 2) and in this way the readers of the website were able to get more familiar with our work and to fill out our survey.

The article contained questions regarding our project, how we developed the idea to create this software, what is our current progress and what our plans are.

The journalist that did the interview with one of our members (Bojan Kovachki) declared interest to do a follow up story once we finish with the Planspiel.

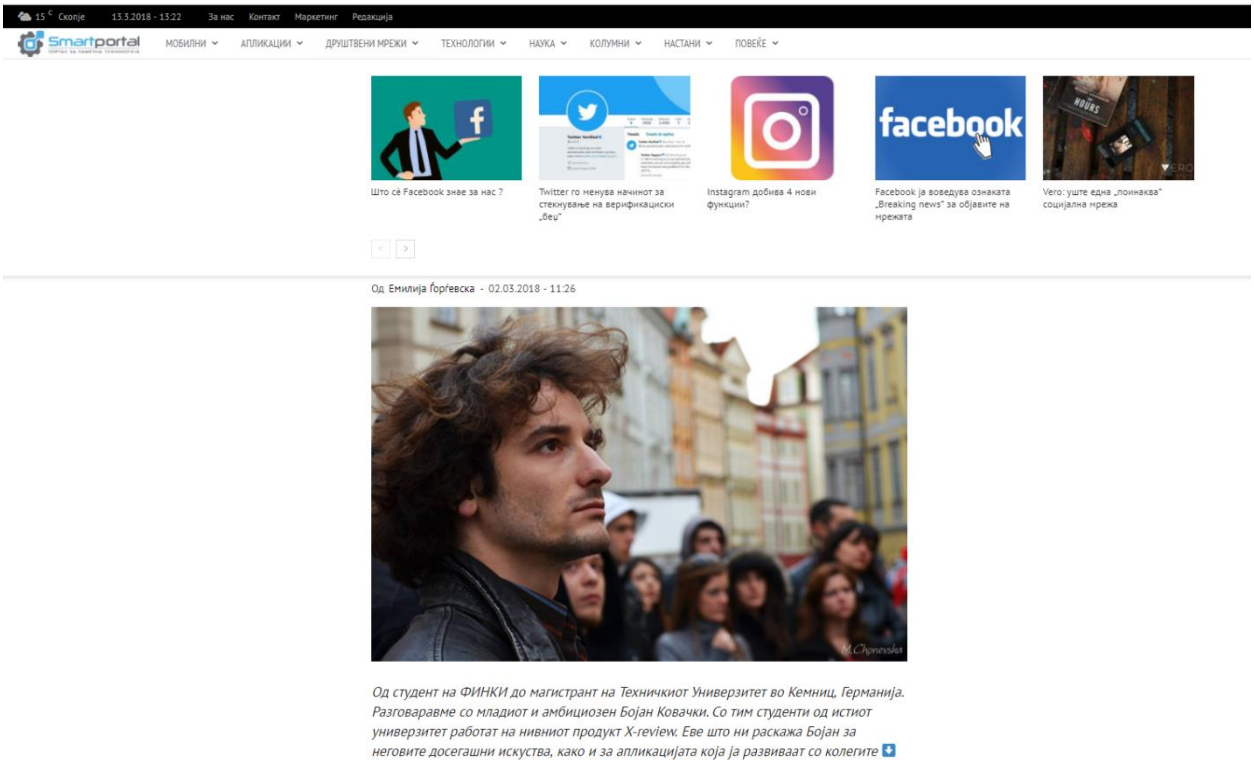

Figure 10: Smartportal.mk Article

Machine Learning with Feature Extraction

We have targeted to analyse predictive price whenever users want to observe that. The system provides us feature extraction from the dataset. In our project, we used real data that has published in a Kaggle contest before [1]. We had a reason to use that real data because the big data system cannot collect so much data in a short time. We also found a research database which has offered by a research institution.

Nowadays, crawling customer reviews and collecting data is an essential part of big data systems. The developer of big data systems should aware of missing values, malformed records, data redundancy or cranky file formats [2]. In machine learning area, the more qualified data you use, the more effective results you get. We have achieved our result with unsupervised learning. Principally, unsupervised learning uses clustering method without finding a hyperplane as the supervised learning did.

In the combination of unsupervised learning and supervised learning (natural language processing), you can create more powerful tool than you utilized the single method. I want to explain briefly unsupervised and supervised machine learning. In supervised learning, you will create a function which takes input parameter to create an output. They call this function as mapping function [4]. Supervised can be grouped into regression and classification

Classification: If you categorize your output with a label, you classify the output. For instance, binary classification, multi-label classification [3].

Regression: It shows the output value as a real value. For instance, Softmax Classifier shows all results based on probability.

Unfortunately, unsupervised learning takes an input but there are corresponding output variables [3]. In unsupervised learning all data is unlabelled. Moreover, unlabelled data usage is a cost-effective way to implement a machine learning algorithm. Our approach depends on unsupervised machine learning and supervised machine learning algorithms together, which calls semi-supervised learning.

Sentiment Analysis and Characteristics Extraction

Tokenization

Sentiment analysis (also known Opinion Mining) is used to extract, transform and analyse data source (text, voice) by using natural language processing, computational linguistics or any other suitable methods [4]. We have started to tokenize many sentences in accordance with positive or negative opinions. At the very beginning, we have used an API [5] to discriminate positive and negative opinions to the language. Furthermore, API implemented a tokenization process on our reviews or sentences. Basically, tokenization is the process of splitting a string into a list of pieces or tokens [6]. Simply, a token is also a piece of a whole, so a sentence is a token in a paragraph [6]. Most of NLP (Natural Language Processing) libraries use supervised learning algorithms, which means that we need to find a lot of labelled data with regards

The tool named NLTK provides many labelled data that you can use easily. We also used this tool in our projects. Our first implementation has been finished with API, but we created sentence tokenizer later.

Part of Speech Tagging

Part of speech tagging is the process of converting a sentence, in the form of a list of words, into a list of tuples, where each tuple is of the form (**word, tag**) [7]. These words can be tagged as verbs, nouns, adverbs, adjective and stop words (the words which are not related to context). Thanks to part of speech tagging, the machine learning system starts understanding the phrases from a sentence. Twitter pre-recorded sentiment analysis collection has been used for positive and negative discrimination in this step.

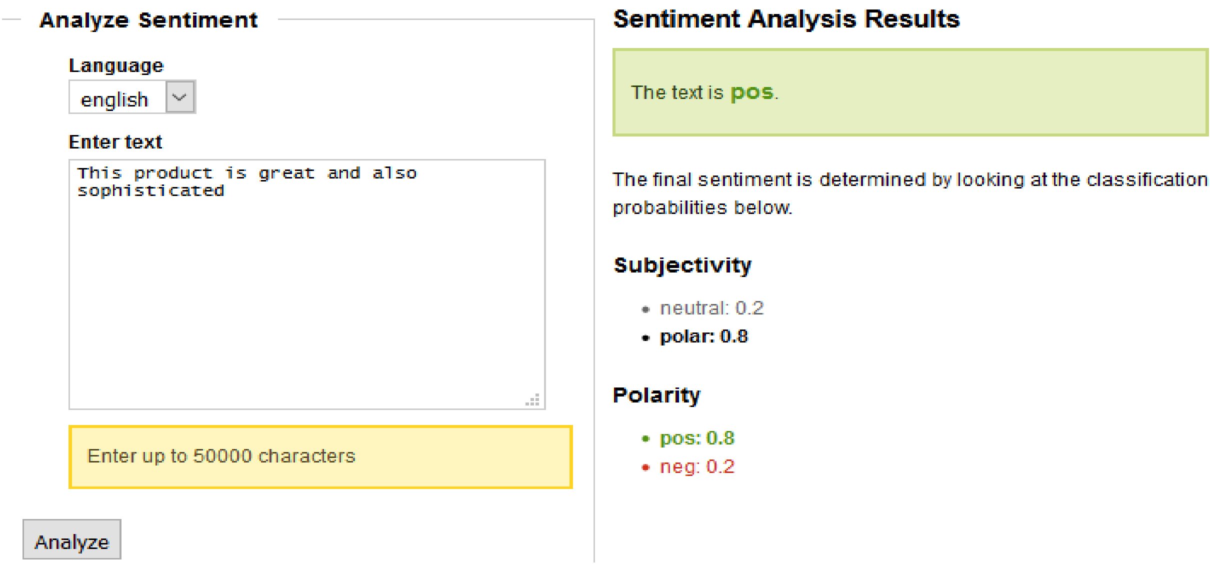

Figure 11: Positive and Negative Sentiment Analysis [5]

In the part of speech tagging, parse tree approach might be used. As shown in Figure 8, the parse tree is the entire structure, starting from S and ending in each of the leaf nodes. (John, hit, the, ball) [8]

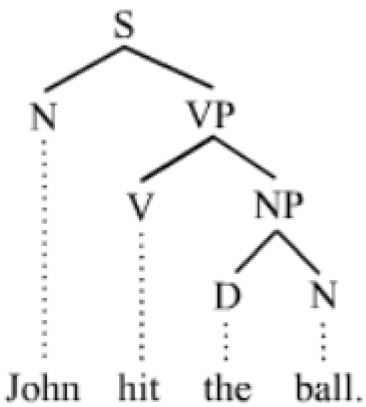

Figure 12: Parse Tree [8]

The abbreviations in the above diagram: S for sentence, NP for noun phrase, VP for verb phrase, D for determiner (definite article - the), V for verb and N for noun [8]. The more complicated sentence structure we have, the more time we need to spend extracting. As you can see, the tree will be unbalanced when we have dealt with complicated and long sentences. Mostly, algorithms will be managed by the library.

Text Classification

Text classification is a way to categorize documents or pieces of text [9]. In this step, we will assign the binary-class or multi-class labels anymore. Classification works by learning from labelled feature sets, or training data, to later classify an unlabelled feature set [10]. After representing all token with features, we can start feature extraction. Chiefly, NLTK uses the bag-of-words method which will be explained in the following statement. For instance, we have these two sentences:

Sentence 1: “Alice is looking for the rabbit hole because she likes rabbit”

Sentence 2: “Bob also likes rabbit, but he does not look for a rabbit hole”

So, every element will be counted according to their frequency.

“Alice”: 1, “is”: 1, “looking”:1, “for”:1, “rabbit”:2, “hole”:1, “because”: 1, “she”: 1, “likes”: 1

“Bob”:1, “also”:1, “likes”:1, “rabbit”:2, “but”:1, “he”:1, “does”:1, “not”:1, “look”:1, “for”:1, “a”:1, “hole”:1

If we extract features from both sentences, we will reach the following result:

(1) [1, 1, 1, 2, 1, 1, 1, 1, 1, 0, 0, 0]
(2) [1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1]

We just extract the features and put an Array List as above. If we find any redundancy of words, we will count plus 1 and save it [11].

Normalization

This method is also called TF-IDF Vector Space. We will calculate the relative frequency and Euclidian Distance between all items. It also normalizes several features that will serve in clustering phase. Let’s think about the following case [12]:

“Alice is my hero” – Document 1

“Bob is not my hero. Alice can be a hero” – Document 2

Document 1

Term	Term Count
Alice	1
Is	1
My	1
Hero	1

Document 2

Term	Term Count

Bob	1
Is	1
Not	1
My	1
Hero	2
Alice	1
Can	1
Be	1
A	1

IDF (“Hero”,document1) = 1 / 4 = 0.25

IDF (“Hero”, document2) = 2 / 10 = 0.5

IDF is constant per corpus (context), and accounts for the ratio of documents that include the word “Hero” [12].

IDF (“Hero”, D) = log 10 based (2 / 2) = 0

Therefore,

TFxIDF (“Hero”, d1) = TF(“Hero”, d1) x IDF(“Hero”, D) = 0 x 0.25 = 0

TFxIDF (“Hero”, d2) = TF(“Hero”, d2) x IDF(“Hero”, D) = 0 x 0.5 = 0

Then we need to eliminate non-informative words just like “Hero” [12].

TF * IDF weightings used for product name standardization as well. When we fetch data via web crawler, product name could be considered non-informative. For instance, “iPhone 5s 16 GB Blue Unlocked”. This weightings method can be used to standardize product name [13]. Raw term frequency suffers from critical problems: all terms are considered equally important when it comes to product name standardization [14]. So, we need to use TF*IDF values for the standardization because the most common words such as “unlocked”, “black” or “dual-core” should be avoided (they have low IDF scores) [13].

As depicted in Figure 9, developers mainly pass three stages if they want to implement a backend software which produces a characteristics extraction. We need to filter out unnecessary data and use clean data to reach a better result. Filtering is an essential step for feature extraction.

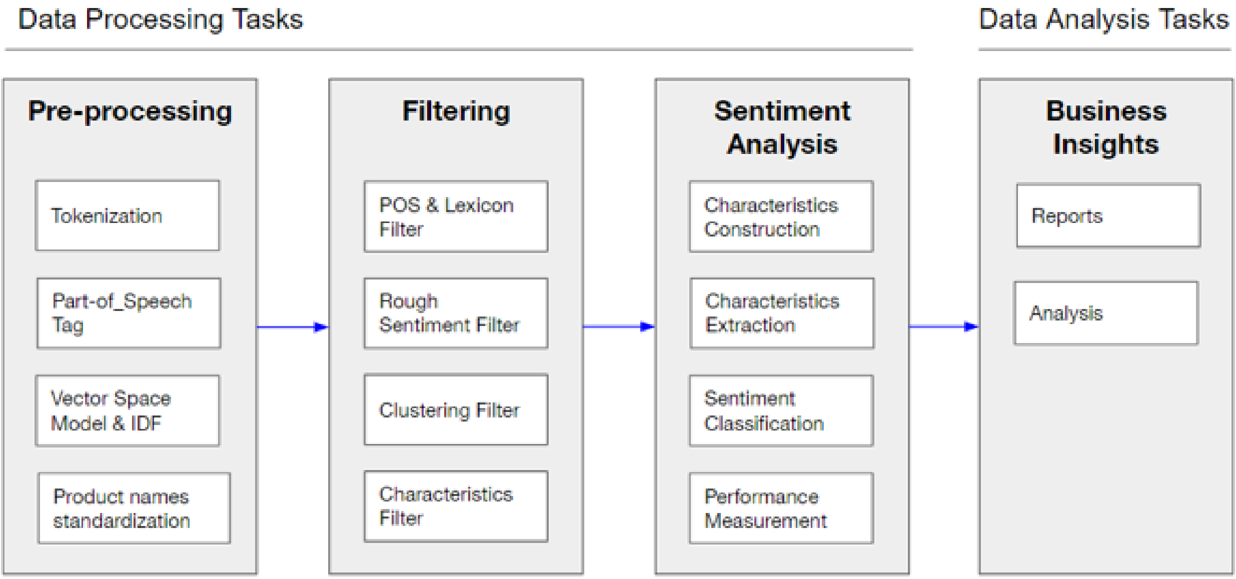

Figure 13: Data Processing Task [15]

POS Filtering: We have tagged all token words in Part of Speech Tagging. Large features need to be eliminated for the final phase so that we take into consideration of keeping nouns, verbs and adjectives [16]. Another part of tokenization will be ignored.

Rough Sentiment Analysis Filter: This filter evaluates the reviews with rating score. The reviews examine for whether their positive scores or negative scores. In case the negative and positive scores have a higher value than predefined boundary values, regarding reviews are to be filtered out. After this step, last reviews which have passed through the filter would assess by rating scores and filtered out once again [13].

Clustering Filter: TDM (TF * IDF) value will be filtered out according to frequency results of bag words.

Characteristics Filter: This filter will eliminate characteristics value of products from predefined dictionary values. After finding a characteristics value with scoring (TDM – TF * IDF), it will diminish the values with regards to characteristics, rating, reviews and scoring [13].

Characteristics Sentiment Extraction

K-Means Algorithm

As we stated, contrary to supervised learning, there will be no label on unsupervised learning. “Clustering is a form of unsupervised learning whereby a set of observations is partitioned into natural groupings or clusters of patterns in such a way that the measure of similarity between any pair of observations assigned to each cluster minimizes a specific cost function” [17].

Figure 14: Clustering [17]

Every cluster needs to have the centre of a mass node. When the centre of a mass node has placed, we can start clustering phase. Therefore, the key parameter in clustering is optimization. Since K-means algorithm minimizes the distortion function, optimization must be implemented in every step [18].

Distortion function shows as below.

Minimize (cluster center, encoder) = Sigma (i=1, N) (|| observations to clustered – cluster center||) ^ 2 [19]

“Product features were those extracted in the unsupervised learning. Feature values that were assigned to each product feature were the k mean distances (the number of words between them) between the product feature and its nearest opinion word (the center of mass) in that document or sentence” [20]. Based on the hypothesis that product features will be assigned in accordance with positive or negative values on a specific test set.

Machine Learning with Price Analysis

Introduction

As we stated in our project description, we have demonstrated a tool which provides us an analysis of prices from different E-Commerce sources such as eBay, Amazon. Price Data

with a timestamp is hard to collect to assess retrospectively. So, we had to simulated price data and we decided to use random price data with a specific timeline.

In Economics, they refer kind of series as time series and “TS (Time Series) is a collection of data points collected at constant time intervals” [21]. Time-price model needs to be stationary series if you want to carry out stochastic methods, which means that the mean of series and variance should not be a function of time.

What does it mean stochastic methods? In statistics, the prediction will define its error.

$$X(t) = X(t-1) + Er(t) \quad (22)$$

The price of specific time interval t depends on the time interval of t-1 and the error which has occurred after prediction. The formulation as above perfectly fits a method named random walk. Broadly speaking, random walk known as a stochastic process that consists of success and failure steps on it. Let's assume you will guess a movement of a person who is in a room that you can never see inside of it. Therefore, you know his movement probability. He or she can move forward, backward or cross walking with one step [22].

$$X(t) = X(0) + Sum(Er(1), Er(2), Er(3)\ldots.Er(t)) \ [22]$$

In machine learning, we need to minimize iteratively unless the error converges to zero. We have used a statistical method for training and a supervised method which called LSTM (Long Short-Term Memory).

The Method of Price Prediction

We have used ARIMA method for training and Deep Learning Method for the test phase. ARIMA method is a statistical method that can be used with non-stationary time series. Time series provide the opportunity to forecast future values. Based on previous values, time series used for the forecast in economics as we implemented [23].

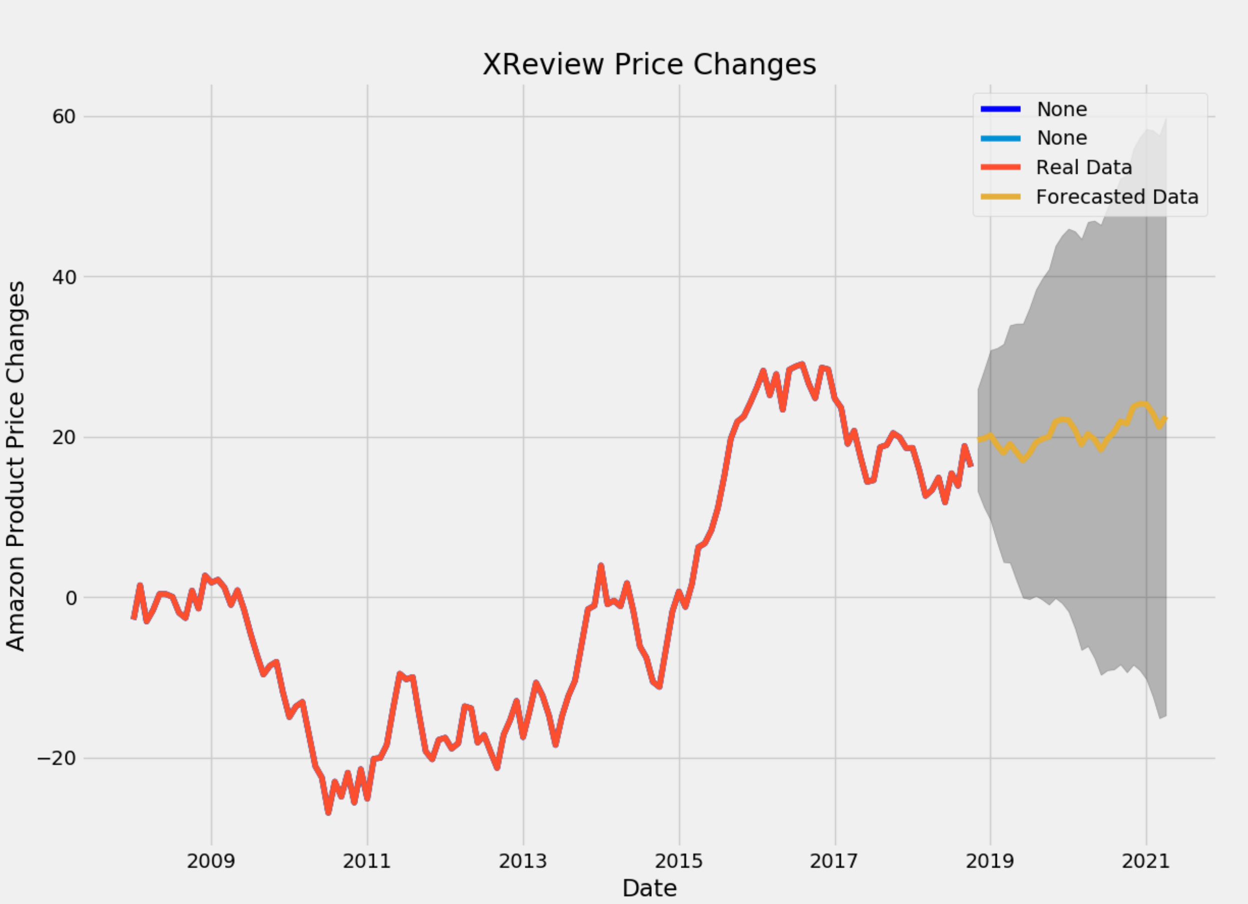

Figure 15: Simulated Random Data [24]

ARIMA stands for autoregressive Integrated Moving Average models. "It is a forecasting technique that projects the future values of a series based entirely on its own inertia" [25]. The data created with a timeline between 2009 and 2011. Data represents a deviation of prices and prediction will show us how much deviation possibly occurs after 2019. This method at least 50 and preferably 100 observations should be available to demonstrate a proper model. The first point that we need to handle is converting series from non-stationary to stationary. ARIMA takes three values which are (p, d, q).

P (Autoregressive Part): Past values changes present values in statistical models. Autoregressive part represents the effect of values. For instance, "this would be similar stating that it is likely to be warm tomorrow it has been warm the past 3 days" [23]

D (Integrated Part): This term shows us the number of differencing in time series lags [23]. If the price-time series that we have created has a small amount of changing between timeline, the predictive values will slightly change on a small scale.

Q (Moving Average): This value represents our cumulative error rate. Error rate increases because the model will depend on cumulative steps. So, we need to decrease error rate each step [23].

Although models need large data requirements, it works for a short run. It has the advantage of being less sensitive to the underlying assumptions of the nature of the data fluctuations [26]. We evaluated a dataset from an academic research for price analysis as well [27].

```
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
    5     26     34      1     0     0   2.254D-04   2.082D+00
  F =   2.0816046902631440

CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH

 Cauchy                time 0.000E+00 seconds.
 Subspace minimization time 0.000E+00 seconds.
 Line search           time 0.000E+00 seconds.

 Total User time 0.000E+00 seconds.
```

{"1201824000000":1.4821627897,"1204329600000":-3.0311807539,"1207008000000":-1.6179703515,"1209600000000":0.3824385492,"1212278400000":0.
{"1541030400000":19.5988725989,"1543622400000":19.7717287071,"1546300800000":20.2075039343,"1548979200000":18.9692064214,"1551398400000":

Figure 16: A sample of ARIMA process [24]

After we acquired ARIMA process result, we send the data via Flask API to Dashboard with JSON format as can be seen in Figure 12.

Testing with LSTM (Deep Learning)

LSTM is a type of recurrent neural network (RNN). “A recurrent neural network (RNN) uses it previous output as an additional input” [28]. Deep Neural Network is one of the best supervised learning methods for feature extraction, but it suffers from Vanishing Gradient Problem.

Basically, the error of cost function should minimize to teach a system in an effective manner. But the error signal of the function starts disappearing toward the back by affecting from the output node (As depicted Figure 12, red node on the right). In that case, there will be no error in input nodes to update the regarding weights.

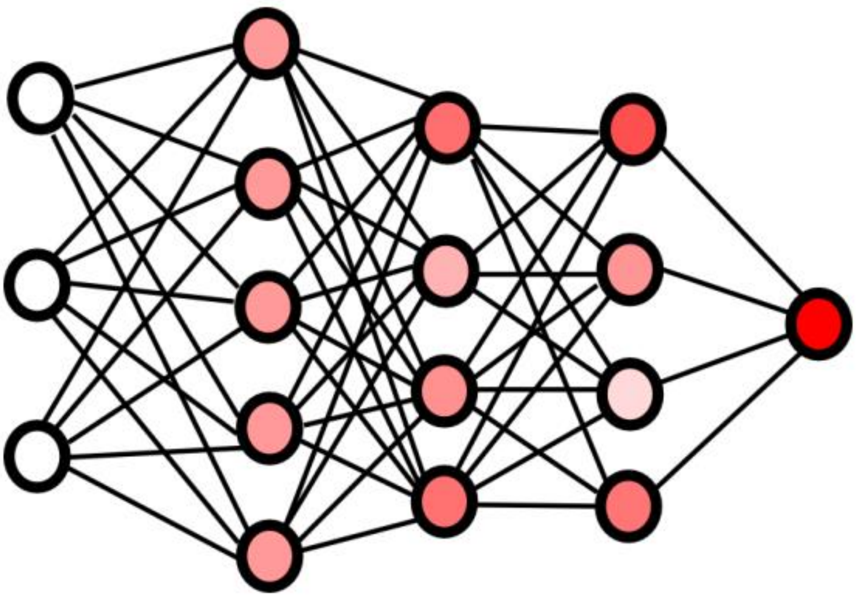

Figure 17: Vanishing Gradient Problem [29]

LSTM solves vanishing gradient problem by holding the error of loss function. It has special gates which decide the information should ignore or save.

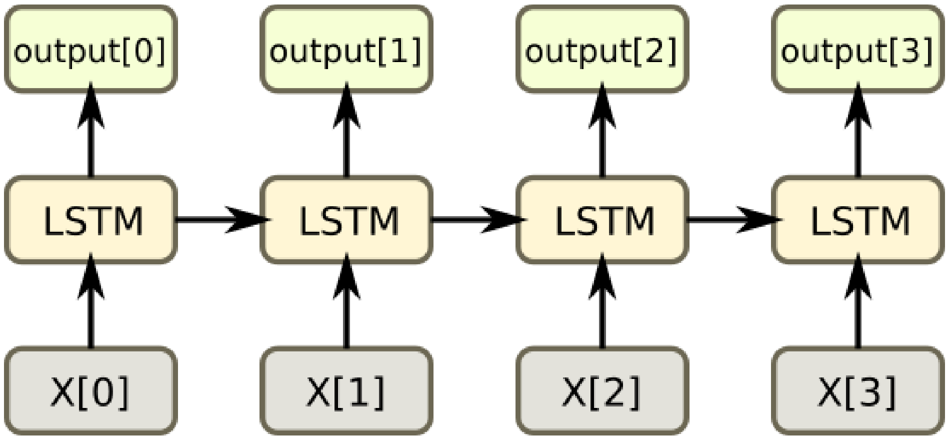

Figure 18: LSTM Network [30]

LSTM catches the error signal between input and output nodes. As shown in Figure 14, LSTM uses one input and produces an output. Next LSTM cell uses the previous output to control the memorization mechanism. When a state has changed in LSTM cell, one of the gates (input, forget and output gates) will take control to decide what will occur the result of a state. Thus, gradient of the error will not vanish when backpropagating because the state remains constant (As depicted in Figure 15) [31].

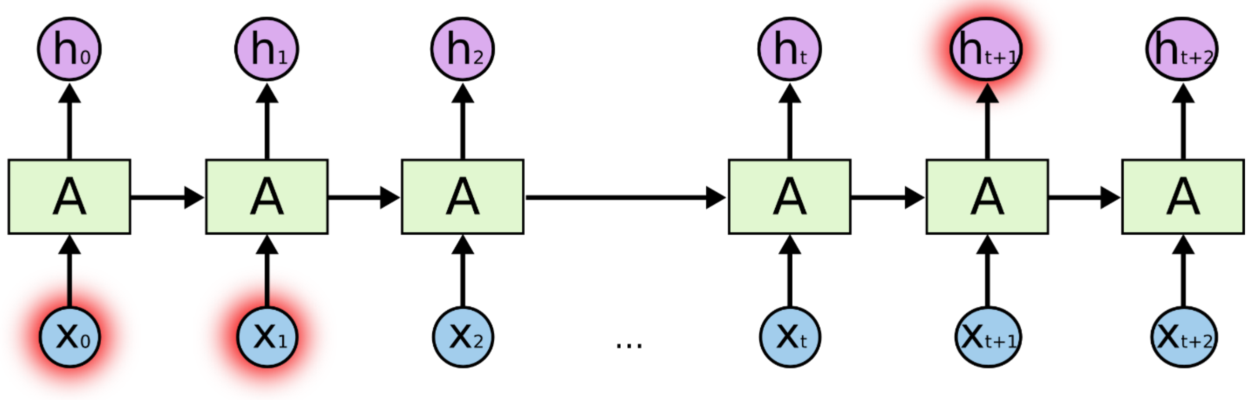

Figure 19: LSTM states remain constant

We have trained price analysis tool 2000 times. By keeping an amount of learning rate as low, we avoid unnecessary convergence. It has used 300 hidden layers. If you use more hidden layer, you can extract complicated features easily. Hence, our case does not need so many hidden neurons.

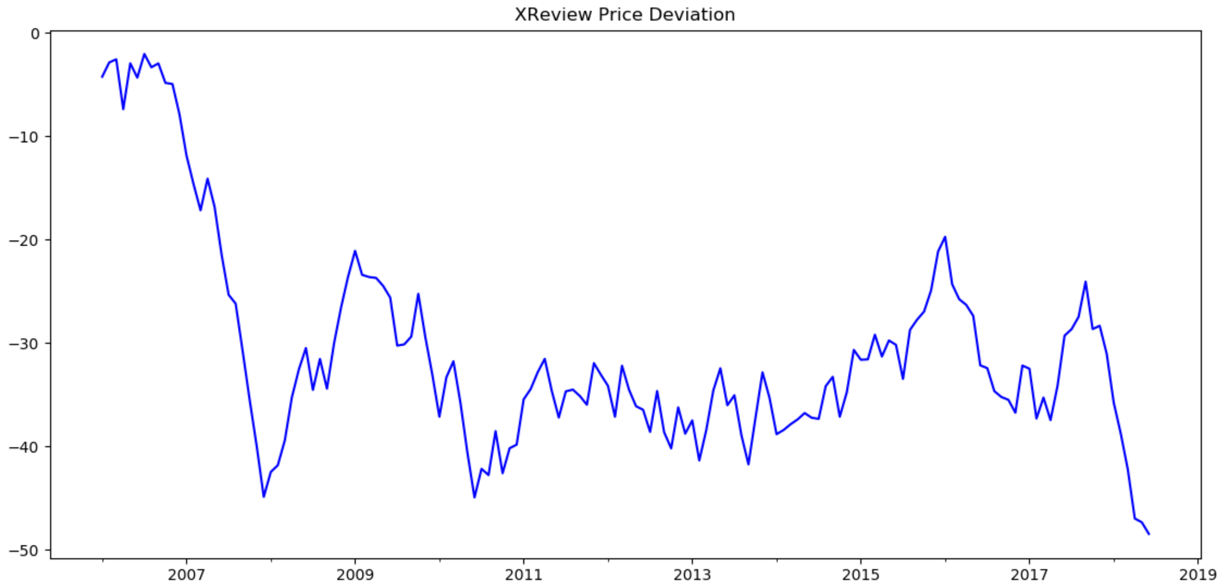

Figure 20: Random data [24]

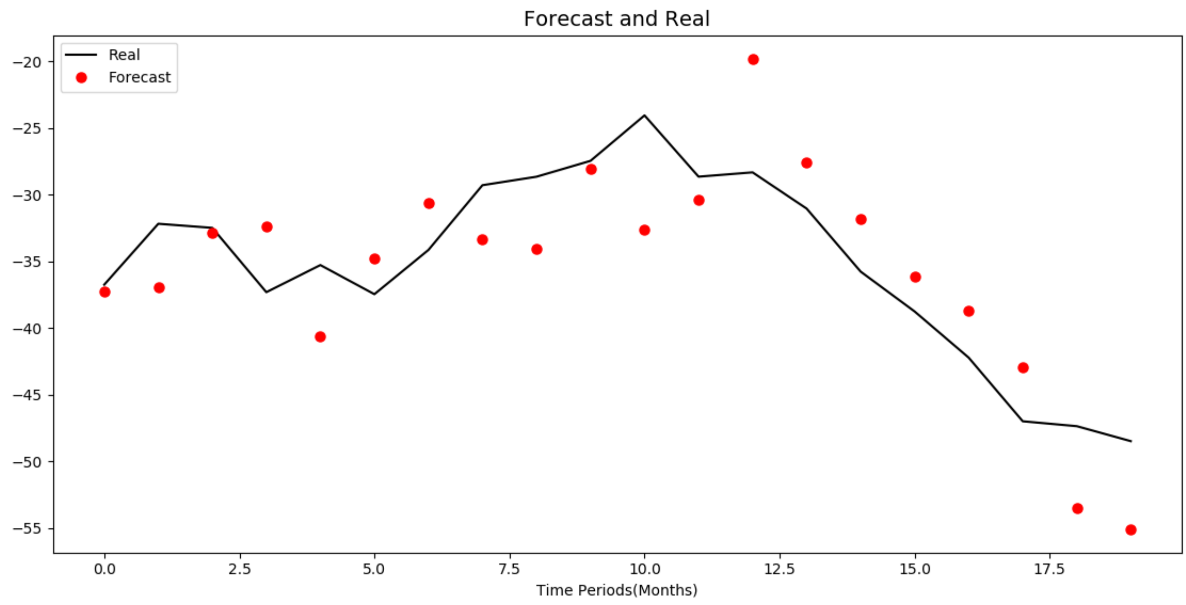

Figure 20: Test of Prediction [24]

Test of Prediction tool has been written in Python and the following libraries have been used such as TensorFlow, Keras, Pandas, Numpy, Theano.

Tensorflow: Google Inc. developed this library for a machine learning algorithm to handle with GPU parallelism. It shows good result in every area like Image Recognition, Text Processing, Complex Feature Extraction

Keras: High level deep learning API which can work on the top of Tensorflow layers.

Theano: It is a Python library that allows you to evaluate mathematical expressions as machine learning algorithm required.

Prediction tool also will inform us the price drop on Amazon or eBay to find the best prices from E-Commerce websites. In Figure 17, the prediction tool demonstrates how much error rate will occur when we run our software. Red points (Forecast data) comply with the black line (Real Data for last two years) without any issues like overfitting or underfitting.

The whole difficulty in Machine Learning is to choose model flexible/complex enough to model the data. If you train your data more than ours, you will reach the boundary of overfitting. The system would learn by heart and this kind of model would not be useful for us. On the other hand, low level of data is not going to be enough for learning phase and it would draw you to underfitting issue (No train).

Magento

Magento is an open source commercial tool that can be used in e-commerce area. Magento is a full-fledged, open source platform aimed at web site designers, developers and business owners who are looking for a complete E-Commerce web site solution [32]. Developers that developed to Magento Software has used Zend Framework (PHP) and the framework allows the separation of Model – View – Controller [32]. Magento aimed two audiences: web developers and business people who related to E-Commerce area. The person who wants to integrate his/her system into Magento can easily customize. The following statements about the advantages of Magento:

All your price and customer data can be control from a single location. There is no longer necessity for multiple login from many locations to handle web sites [33].

Magento has built-in web service. Developers can use SOAP and XML-RPC protocols in an effective way.

Magento has state-of-art reporting feature to the owner of business. You can export the information in a CSV format to integrate with Excel and other database programs [33].

Magento parts use loose-coupling technique so that it allows for easy updating of images and CSS without affecting core functionality [33].

Moreover, Magento has a bunch of themes which you can use in your E-Commerce advertisement e.g. Logo, Shopping Cart, Login and Account Management, Organization, Planning and Workflows.

Logo: It is an essential feature that lets a customer know exactly what kind of a store they are viewing [34].

Shopping Cart: In E-Commerce Store, it is a very important feature to give permanent opportunity customers [34].

Login and Account Management: Customers need the ability to manage accounts and view previous orders [34].

Planning: Your products need to be organized into categorized as shown in Figure 18. The layout also should be aligned with regards to applicability [34].

Figure 21: Magento Demo Store [34]

BlueFoot Idea

BlueFoot is a page builder and CMS solution which integrated with Magento Open Source Software (Version 2.0). It is a powerful set of tools to help you create and manage your content of pages which incorporated with Magento Store [35].

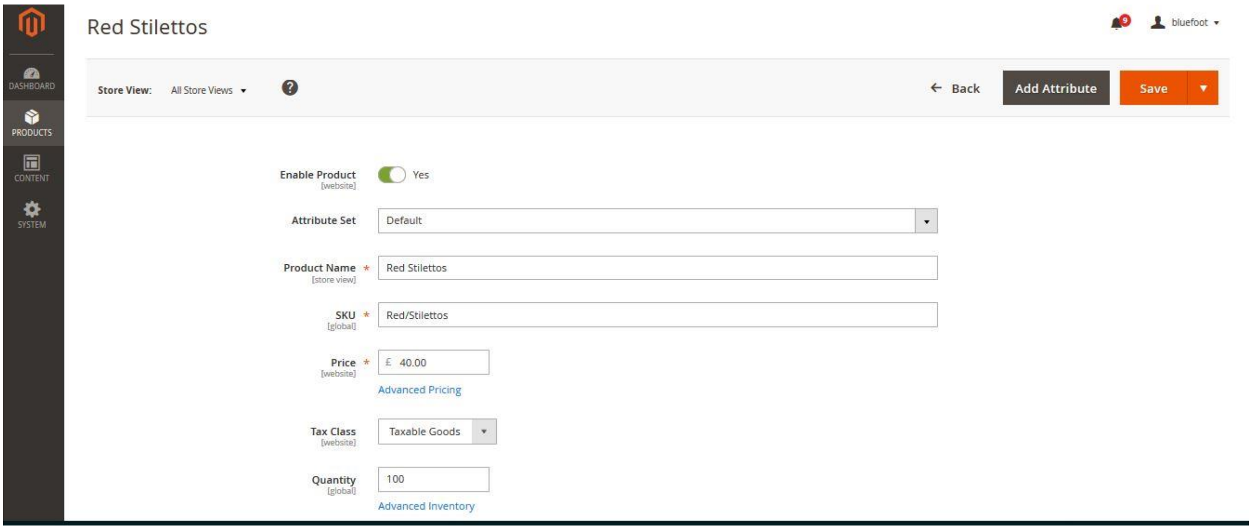

Figure 22: Store View of BlueFoot [36]

We want to apply a product follower from BlueFoot Dashboard which has assigned by the owner of dashboard. In Figure 19, BlueFoot has a feature called Advanced Pricing. A business owner can give a campaign for special products by decreasing the price of a specific product.

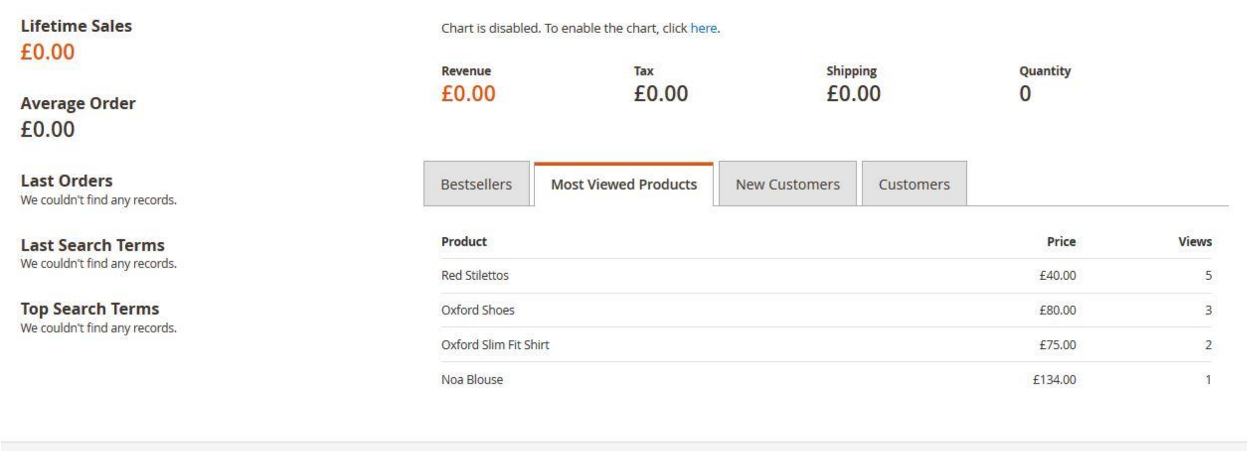

Figure 23: Most Viewed Products [36]

As you can see in Figure 20, the Dashboard shows us most viewed products and customer information. In case of fetching a historical data analysis from Amazon or eBay, the owner of a business may have an idea which product prices should change to increase the rate of sale. We believe that our plugin achieves remarkable results in many products, especially in E-Commerce area.

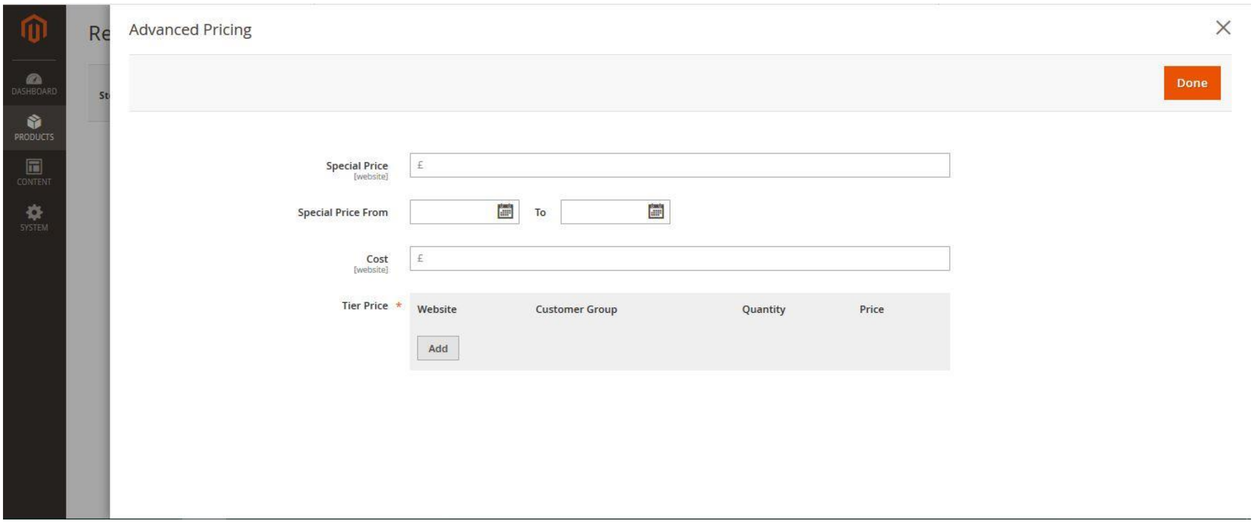

Figure 24: Special Price [36]

Architecture of BlueFoot Idea

The plugin will connect from XReview Software as depicted Figure 22. When a user signs in our software, he would choose a tab named module which consists of Magento plugin. A plugin doesn't touch any of core functionality both sides and it will detect recent updates automatically. When a specific product which resides in BlueFoot Dashboard of XReview, we will show them regarding price analysis. Otherwise, we will start following the product to create a price analysis.

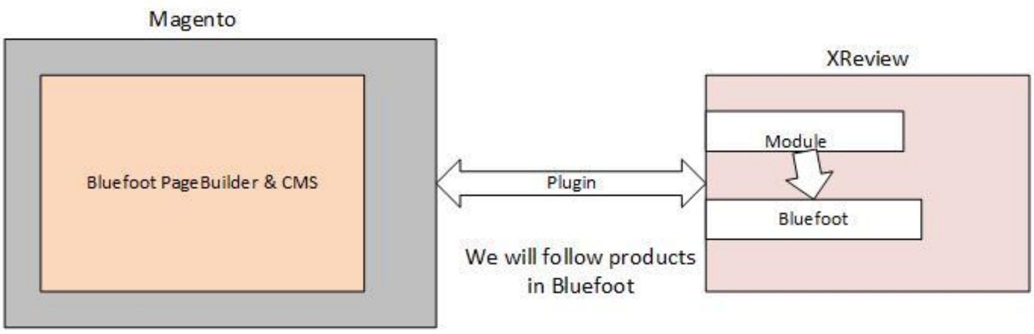

Figure 25: BlueFoot Plugin [36]

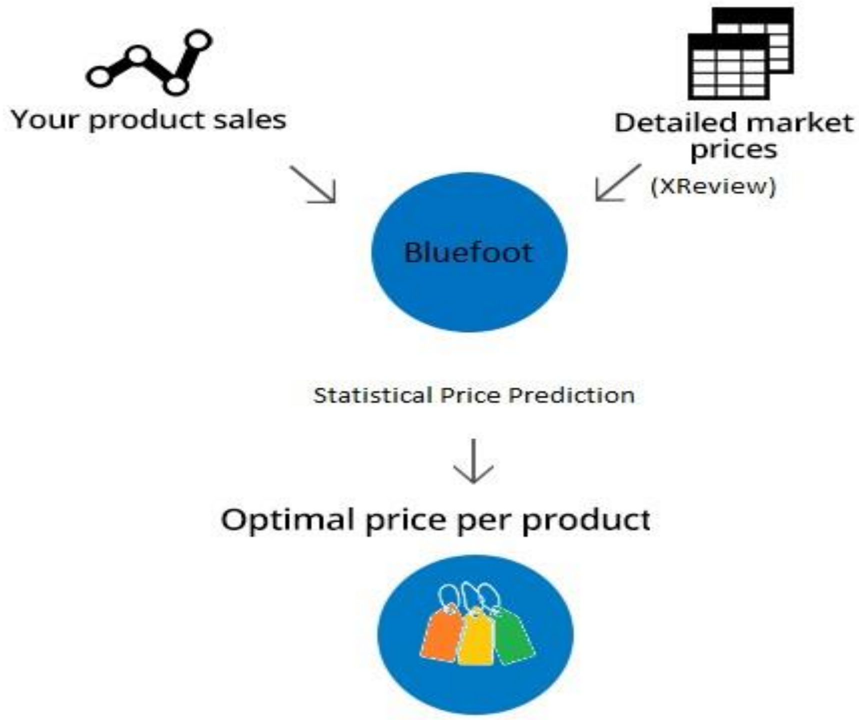

Figure 26: Optimal Price Concept [40]

As depicted Figure 23, XReview will collect the price analysis which a product in the product sales list of any company. At the stage of Statistic Price Prediction, the system will give us the best price information in the past or register into its databases. A business owner can take price data with lowest and highest limit which happened over the past few years.

Future Development

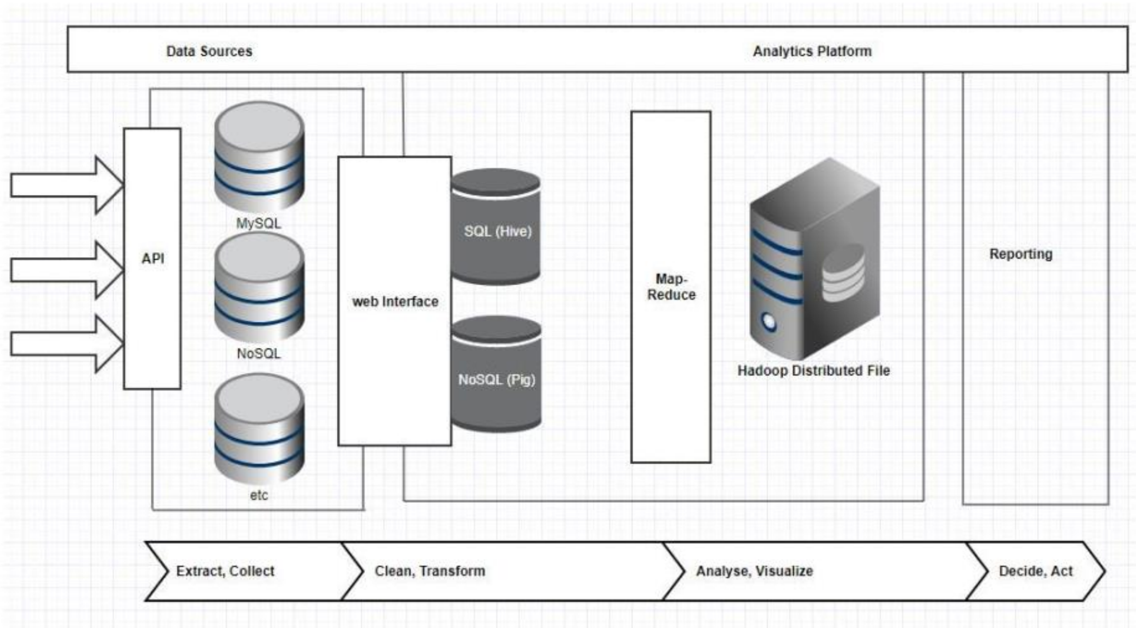

Figure 27: Scalable Architecture [38]

If we may crawl terabyte dataset from E-Commerce, we need a scalable system as above. First, we need to fetch all data via an API and store the data into databases. Our research shows us that we can use effectively relational and non-relational data together. Databases need to comply with the current technologies in Big Data Area like Hive and Pig which are affiliates with Apache. Distributed File System must be integrated to achieve most effective result when we planted the distributed nodes.

Network nodes could be in a failure where they have placed. Principally, you need to use a different system which can resist fault-tolerance. For the common case, data replication will be supported by our future development. When we get a terabyte of data into our network, we need to balance our cluster as well. Hadoop and MapReduce technologies can handle this issue as efficiently as possible [37].

Therefore, if we would pass the stages of Extract, Transform, analyse (ETL) successfully, we have clean data to provide Business Intelligence software. It is not only that we need to create a report which consists of basic assessment, but also, we may offer data integrity in case of data corruption [37].

XReview Architecture

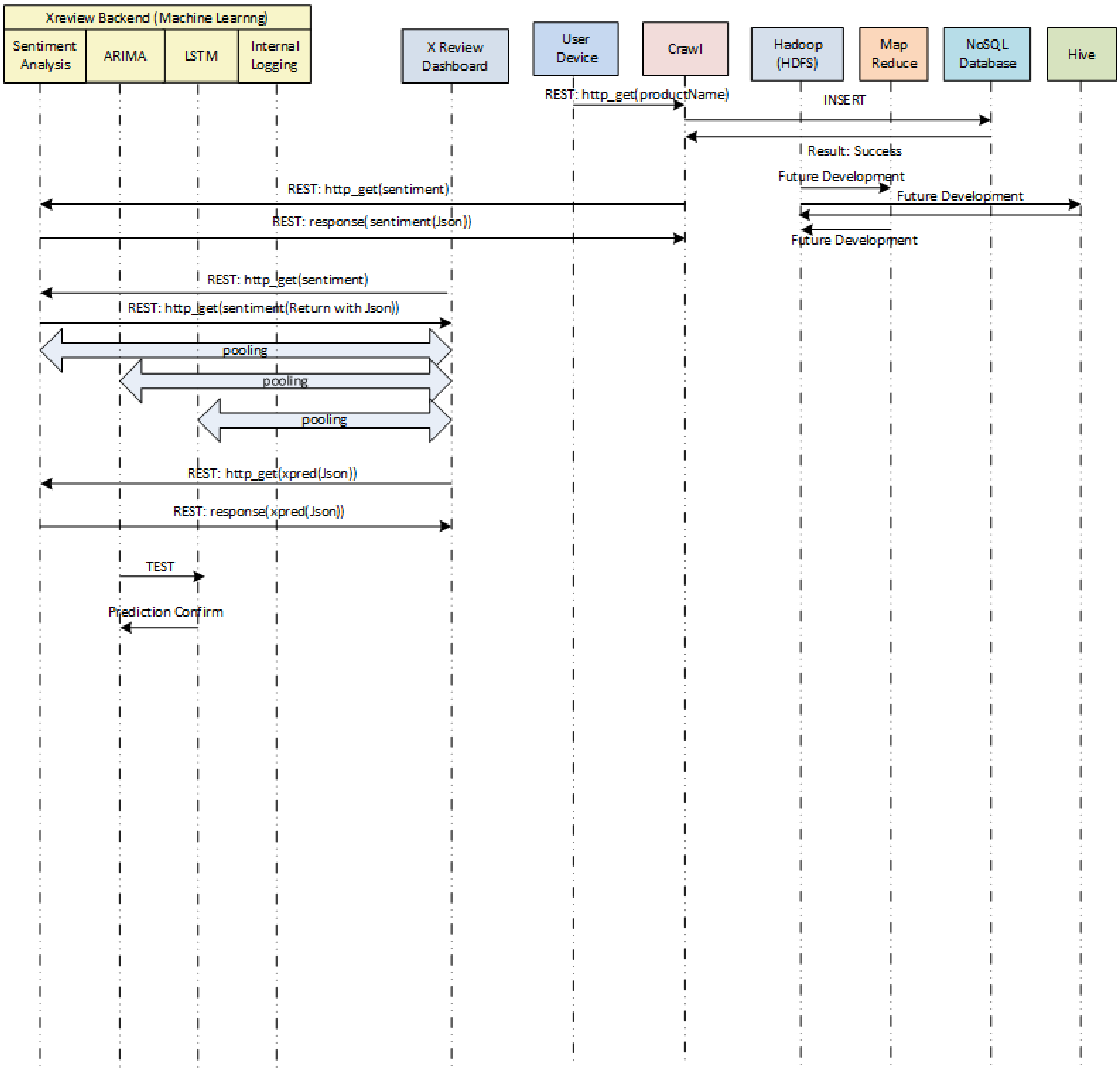

Figure 28: XReview Architecture

Spring Boot and AngularJS

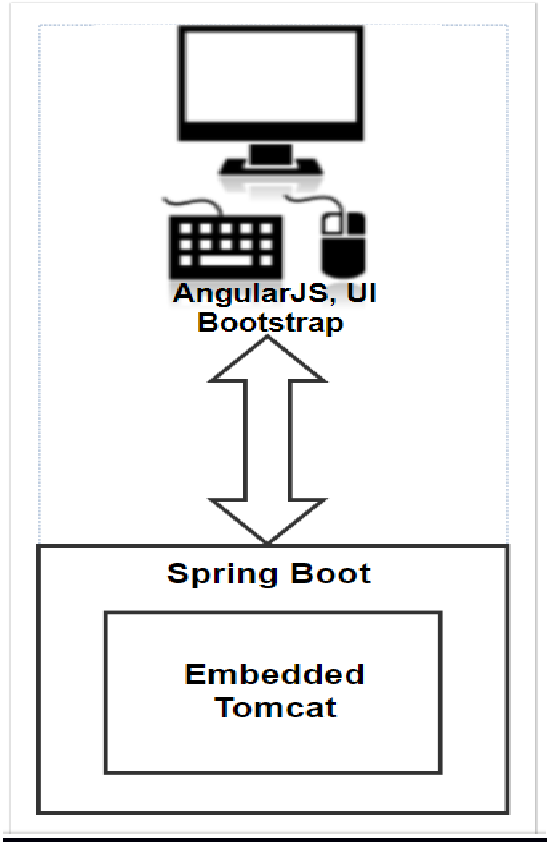

Figure 29: Angular JS and Spring Boot

Spring Boot makes it easy to create stand-alone, production-grade Spring based Applications that you can "just run". We take an opinionated view of the Spring platform and third-party libraries, so you can get started with minimum fuss. Most Spring Boot applications need very little Spring configuration. Spring framework provides flexibility to configure beans in multiple ways such as XML, Annotations, and JavaConfig. With the number of features increased the complexity also gets increased and configuring Spring applications becomes tedious and error-prone. The Spring team created Spring Boot to address the complexity of configuration.

Combining Angular and Spring Boot is a great way of getting up and running quickly with a new web application. However, it was challenging to fit all the different pieces together in the beginning. Our application is UI responsive for mobile devices, uses a data layer with Spring data, and built client-side functionality with Angular. We generated the Angular2 application using angular-cli. This allows us to easily generate a new project with a clear structure. This will also make it easy to add new elements to our Angular2 application. When adding new elements with angular-cli, we maintain the same structure and wire the new elements together automatically. Angular-cli is an npm module, so it requires Node and npm to install and run. Make sure to install a recent version of both applications before continuing with the steps below. We have used Maven as the overall build manager for our application. We define all dependencies of the project in pom.xml.

This gives us a full development environment designed for building large, highly Agile applications is complex and must support the various needs of multiple teams working on the same project. The frontend and UX teams work on the client-side of the project to develop, test, build, and deploy it to a web server (or cluster of web servers); while the server-side team implements the APIs, and deploys them to a scalable environment (SOA or Micro Services). With greater complexity comes a greater cost. Most organizations have some form of folder structure based on predefined (and sometimes arbitrary) corporate perspectives to organize their code bases. This folder structure is different from the functional structure of modules (for client side) and packages (for Java Server side). The prerequisites for our project are listed below:

- JDK 1.8
- Eclipse Mars IDE
- Angular-cli
- Angular 4
- MySQL 5.7
- Npm 5.6.0
- Gulp 3.9
- browser-sync 2.1.8
- bower 1.8

Project Structure

The Following project structure for Client Front end is used:

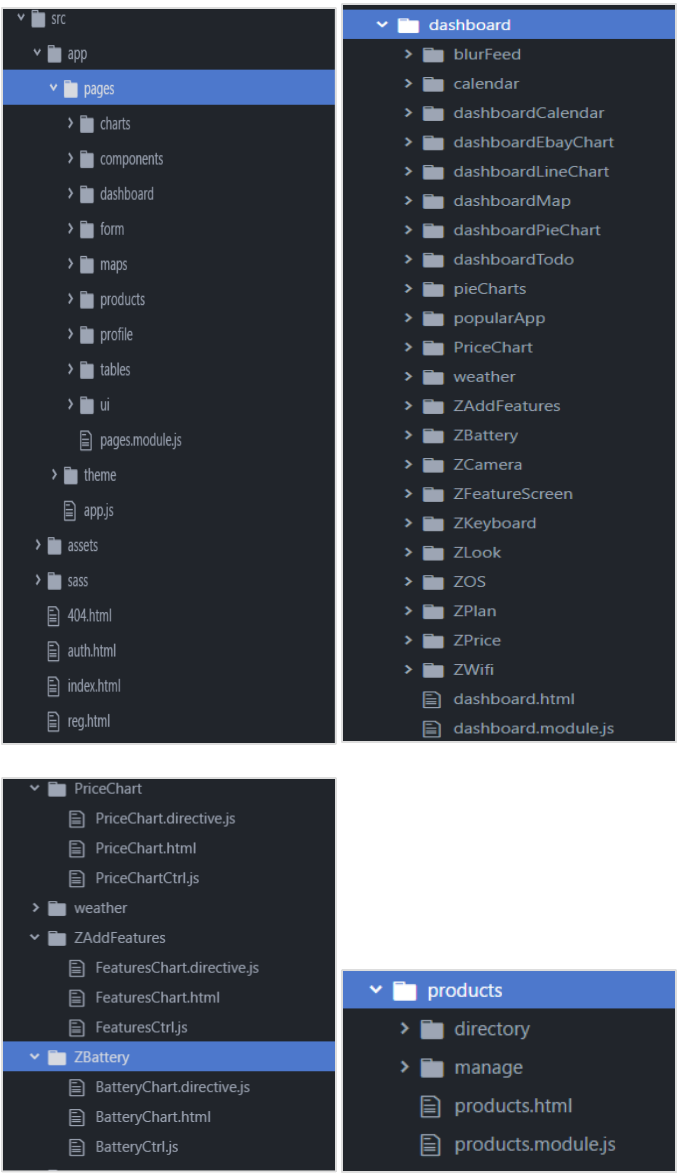

Figure 30: Project UI Folder structure

The application is built using the blur admin template [37]. The app folder contains app.js where the angular module for the entire application is created. We also use ngRoute module, which provides routing and deep linking services and directives for AngularJS apps. We have mapped URL’s to views and added specified controllers to it.

- src/app/pages: It contains all files defined for the application for respective views.
- src/app/pages/dashboard: It contains respective dashboard related files.

- src/app/pages/dashboard/zbattery: It contains all html pages, controllers and directive for battery sentiment of a product. Other sentiments are also added similarly in this folder structure
- src/app/pages/products/: This folder contains the controllers, directives and views for products. Products directory and products add pages are in present in this folder.
- src/app/pages/products/manage: Controllers for adding products are present in this folder.
- Index.html – Angular ng-app is created in this page.

The Node.Js modes are present in the following folder:

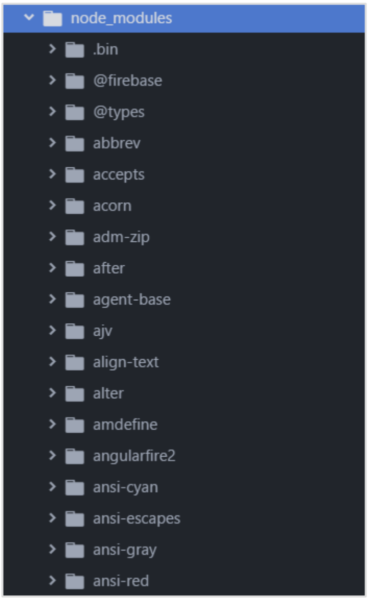

Figure 31: Project Node Modules Folder structure

User Interface

Login page

End user's login to the application using this page. The username and password are to be provided in the respective fields. If the user is new he can click on Sign Up to create a new account.

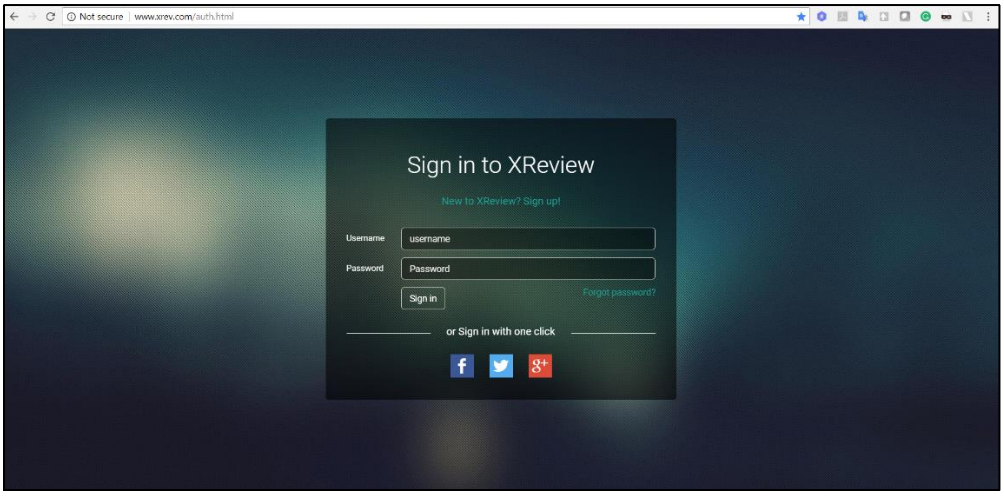

Figure 32. Login page

Sign Up

The new user signing up must provide a unique email id to sign up to the application. The user name and password are provided in the respective fields. After the user signs up, he/she can login to the application with the unique password and user name.

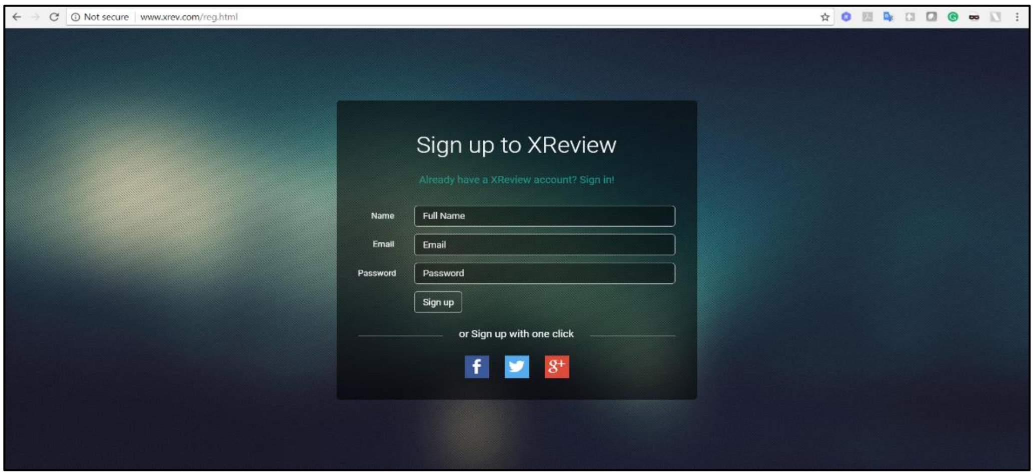

Figure 33. Sign Up page

Dashboard

This will be the second page once the user login to the system. The end user can see the details of the most recent product that he added to the system. The dashboard includes the price comparison from different e-commerce websites and provides the best one for the customer. All the customer reviews and sentiment analysis are also provided here. The predicted price drops along with a time, suggested seller and overall review analysis.

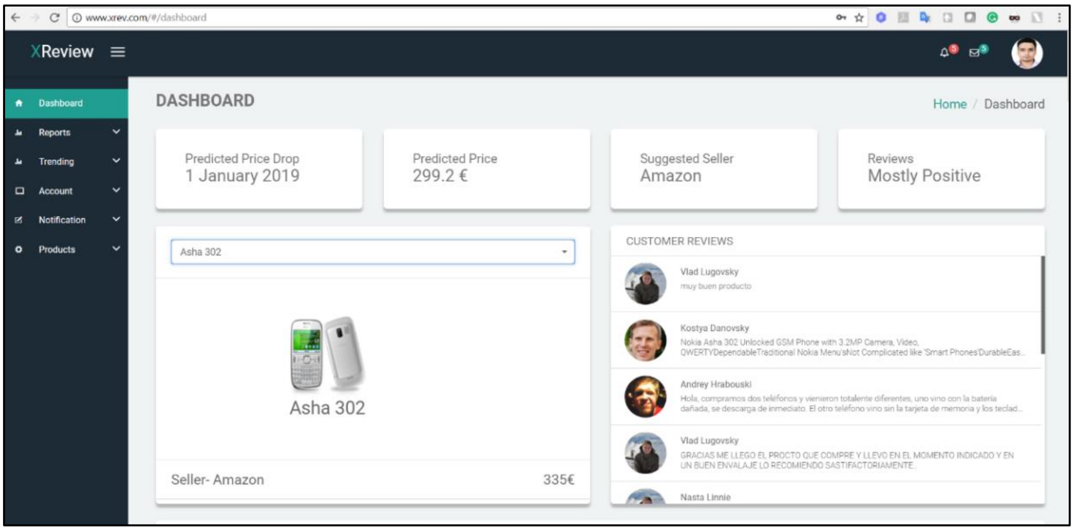

Figure 34. Dashboard

Dashboard Product List

The dashboard contains a search for all the products the user is following. The user can select a product and all the details are displayed on the dashboard.

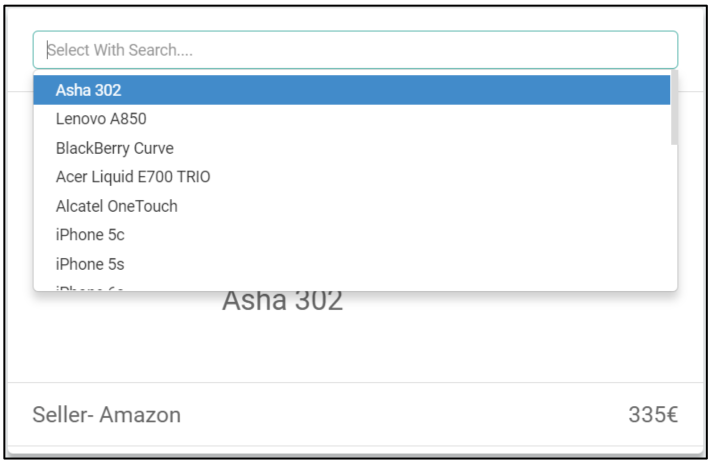

Figure 35. Dashboard product list

Customer reviews

When a product is selected, all the customer reviews collected for the product from a variety of e-commerce websites are displayed for the customer. They are useful to a customer because it’s a concise and summarized form.

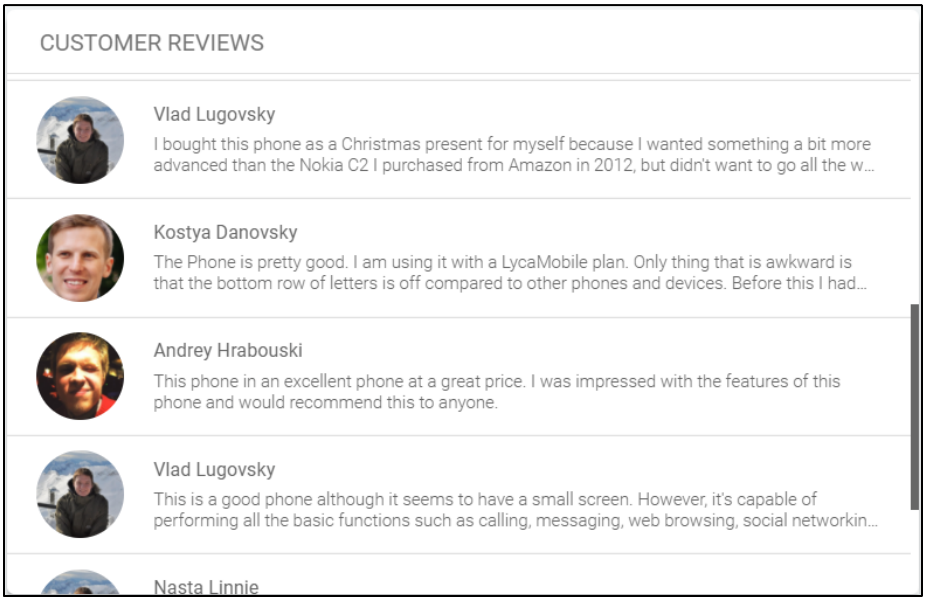

Figure 36. Customer Reviews

Price Deviation

The historical price deviation for a product is displayed, along with the predicted forecasted prices that are calculated from our state of the art machine learning algorithms. As seen in the below diagram, the price drops on Jan 2018 at a price of 299 euro.

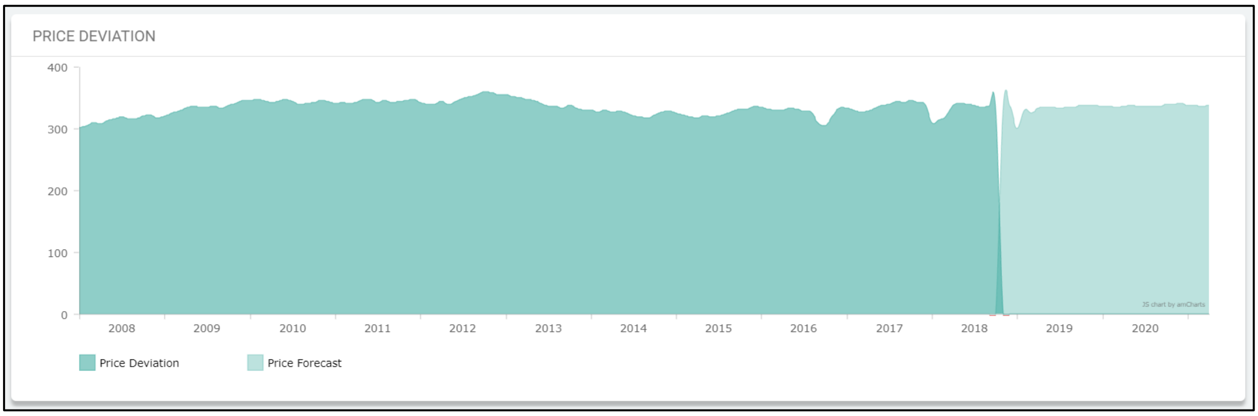

Figure 37. Price Prediction Graph

Customer Sentiment

The dashboard extracts all the sentiments of users from the customer reviews and shows a summarized view of all the sentiments. In the below diagram for the product the user selected the keyboard got overall good response, the WIFI was very good and the screen was poor.

Figure 38. Customer sentiment analysis

Dashboard Suggestions

The suggestions part of the dashboard includes the predicted price drop for the product, along with the predicted price drop. The suggested seller is also provided to help the user decide from which e-commerce website to buy the product from. The suggested seller suggestion is based on the best price and the reviews provided to the sellers from the customers. The overall sentiments of the product are analysed, and outcome is provided to the end user. Here the product has mostly positive reviews.

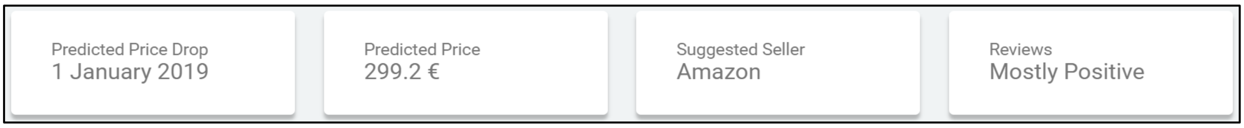

Figure 39. Dashboard Suggestions

Add Products

The add products page asks the user for a product to be followed, the e-commerce websites he wants to buy the product from (This is optional), with help text box asks the user for a product URL (This is optional), and a product description. After the product is being added, our system gathers all information from different websites and analyses them and predicts prices. The user is notified when the process is complete. The product will appear in the main dashboard as well as the product timeline.

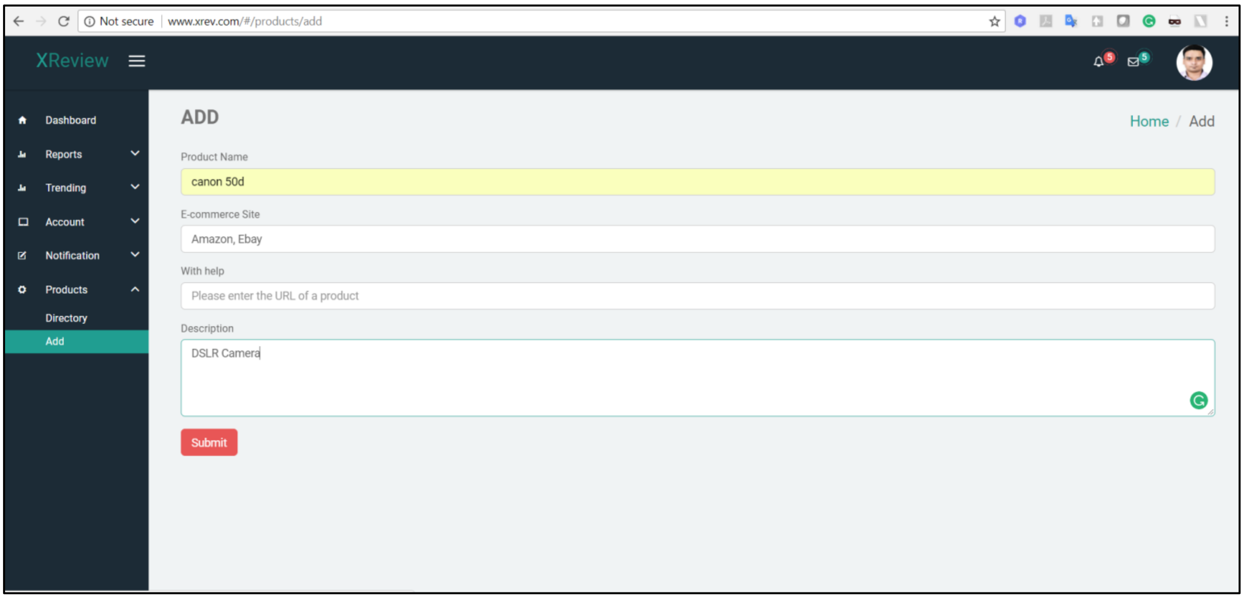

Figure 40. Add Products Page

Products Timeline Directory

The products added to the system can be viewed with the timeline in the directory tab. Here "Asha 302" is the last product that was added to the system on 12 March 2018.

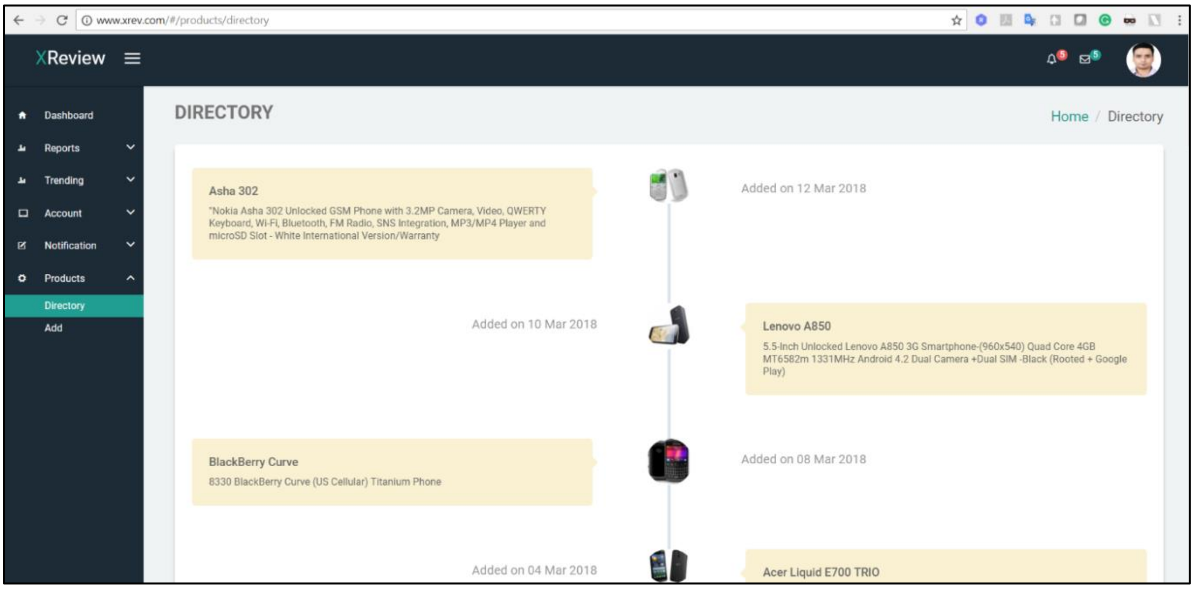

Figure 41. Products Directory

End user Profile

The profile option can be found for the user at the header as shown below:

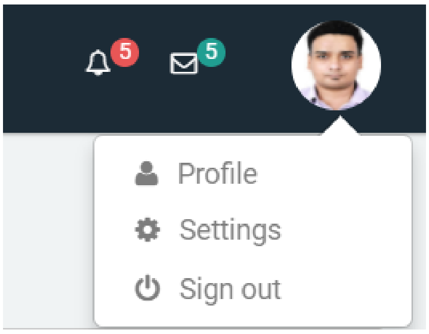

Figure 42. User Profile Option

The profile contains all details of the user. The user should provide a unique email id for the account. The password of the account could be updated form this page.

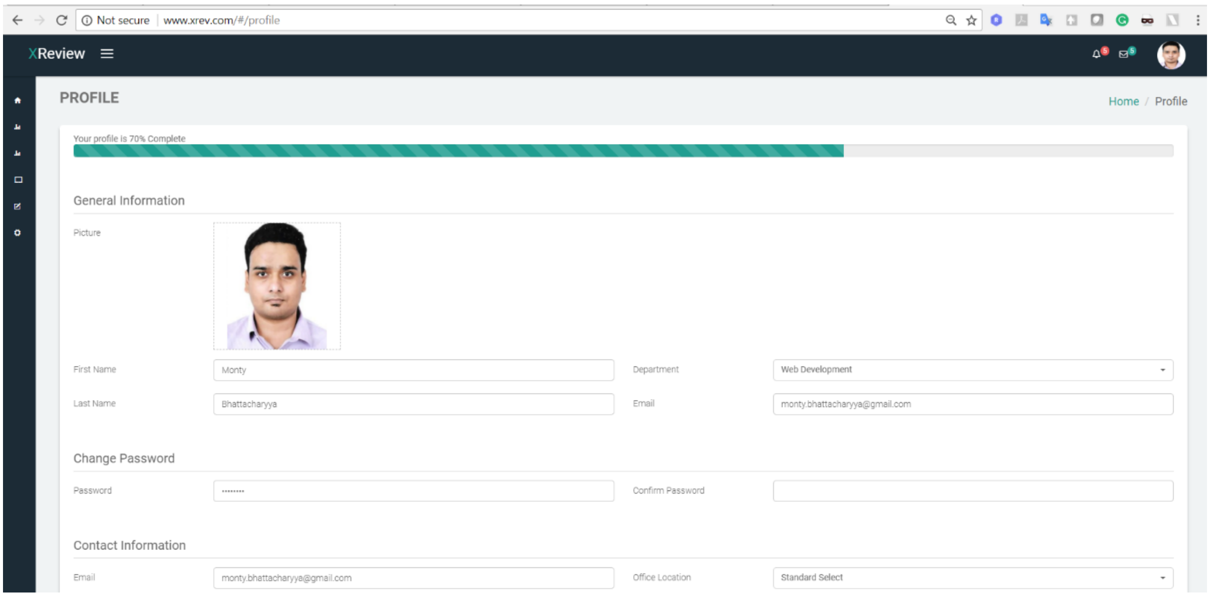

Figure 43. User Profile

The profile also includes the social profile details of the users. For example, here the user has provided Facebook, Twitter and LinkedIn details. This data helps the application to provide a better result. The profile has an option for Email Notifications. The user can set to turn on notifications when Price drops for trending products, when popular products are added, when the products the user is following get negative reviews, when the products user is following gets positive reviews, when new price predictions are available and for daily reports.

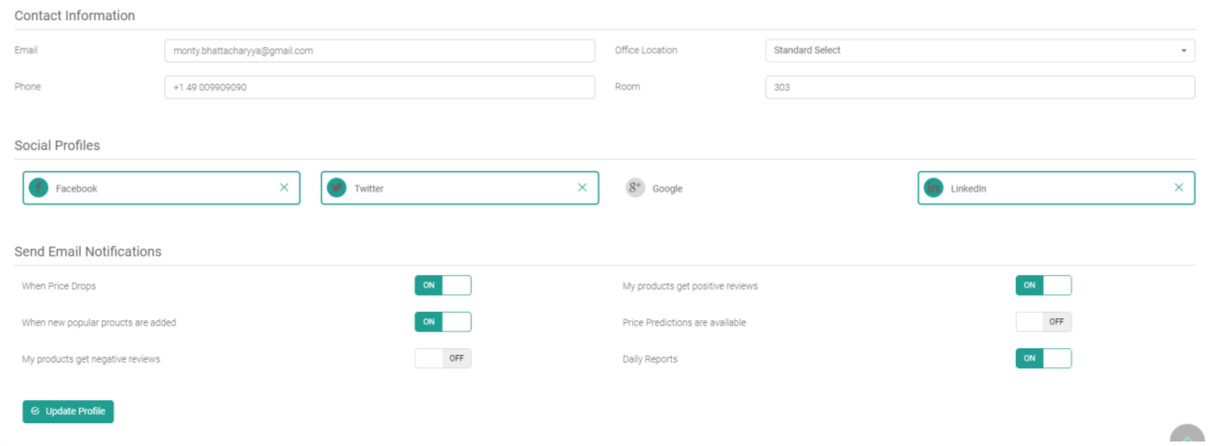

Figure 44. User Notification Option

User Notifications

The Top right menu shows the notifications and the emails that the user has received. In the below example the user received a mail that a new product was added, and a notification that there is a price drop of 10 euro for Blackberry curve.

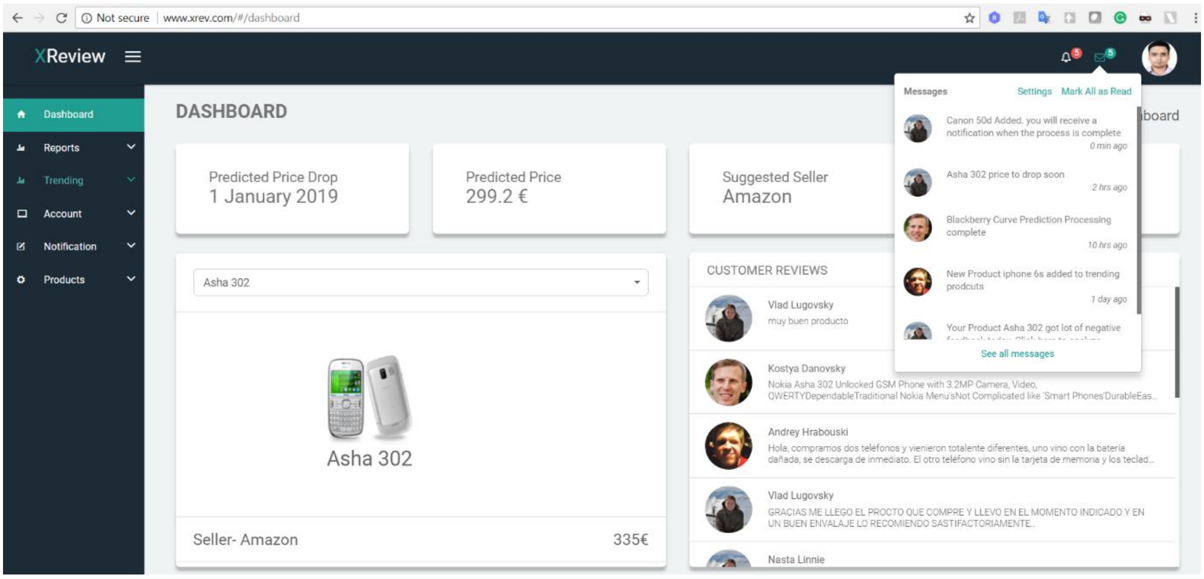

Figure 45. Dashboard Mails

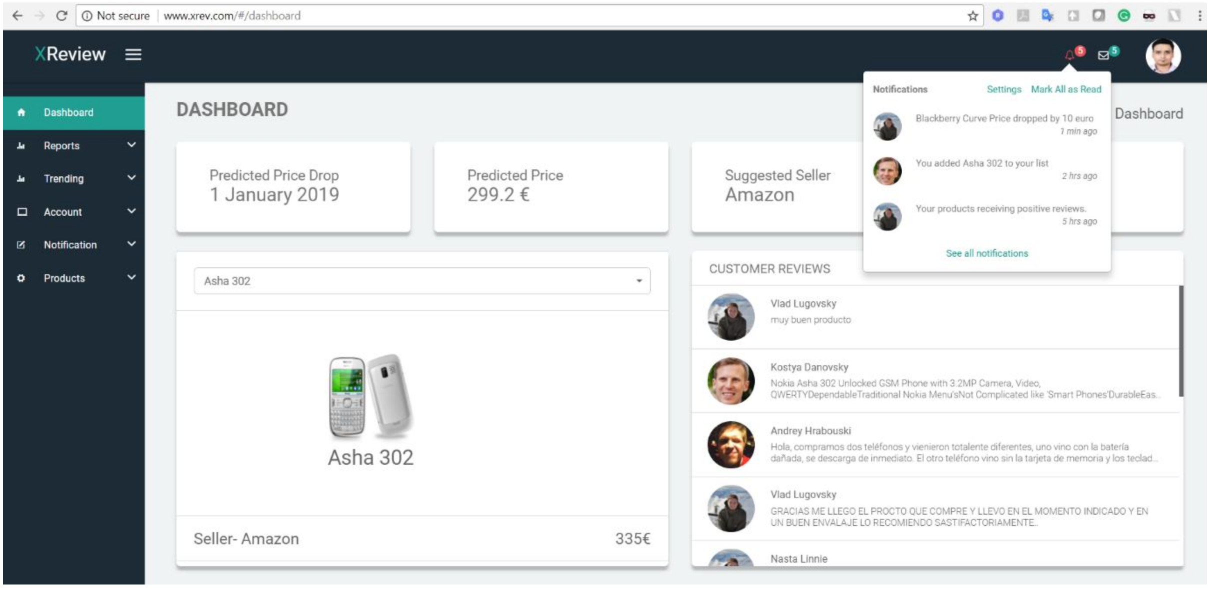

Figure 46. Dashboard Notification

Conclusion

In this report we have explained the features our product called X-Review offers to our customers: customer review analysis, historical price analysis, price drop prediction and price comparison.

We have stated what our company's vision, mission and goals are and have given an overview of the company's summary.

In addition, we have done a thorough market assessment and have described our financial plan in full detail.

Furthermore, our team has created two surveys and managed to get our company's story published in an online article, which helped us reach out to 535 potential customers.

Our motto 'Innovation is our passion!' has been a guiding principle in the development of our product (which has been built using Big Data Analysis and Machine Learning algorithms with Feature Extraction) and its presence is evident in our Magento (BlueFoot) idea.

X-Review is a modern web application that utilizes a minimalistic design. The main accent of the user interface has been put on simplicity and effectiveness and the goal is to make the app as intuitive as possible for the prospective users.

Cutting edge web technologies have been used for development of both the frontend and the backend of the application, as well as the databases.

We believe that with the creation of this product we have totally changed the online shopping experience and have left an everlasting footstep in the e-commerce world. Customers from now on will always reach out to our software before making their final decision on which product to buy, when to buy it and which e-commerce platform to use for buying it.

References

[1] PromptCloud, Kaggle [Online]. Available: https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones/data (26.03.2018)

[2] Q. Ethan McCallum, O'Reilly Press, Bad Data Handbook (pp – 2)

[3] Jason Brownlie [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ (26.03.2018)

[4] Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Sentiment_analysis(26.03.2018)

[5] Wikipedia [Online]. Available: http://text-processing.com/demo/sentiment/ (26.03.2018)

[6] Jacob Perkins. Python 3 Text Processing with NLTK 3 (pp – 8)

[7] Jacob Perkins. Python 3 Text Processing with NLTK 3 (pp – 85)

[8] Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Parse_tree (25.03.2018)

[9] Jacob Perkins. Text Processing with NLTK 3 (pp – 187)

[10] Jacob Perkins. Text Processing with NLTK 3 (pp – 188)

[11] Bag-of-Words, Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Bag-of-words_model (25.03.2018)

[12] TF-IDF Method, Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Tf%E2%80%93idf (25.03.2018)

[13] Ignacio Recasens. Sentiment Analysis, Github [Online]. Available: https://github.com/irecasens/nlp_amazon_reviews/blob/master/Sentiment%20Analysis%20Notebook.ipynb (25.03.2018)

[14] Stanford University. Natural Language Processing, 2008 Cambridge University Press [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/inverse-document-frequency-1.html (25.03.2018)

[15] Ignacio Recasens. Methodology, Github [Online]. Available:

https://github.com/irecasens/nlp_amazon_reviews/blob/master/Methodology.PNG (25.03.2018)

[16] Peter Teufl, Udo Payer, Guenter Lackner. From NLP (Natural Language Processing) to MLP (Machine Learning Processing) [Online]. Available: https://online.tugraz.at/tug_online/voe_main2.getvolltext?pCurrPk=52918 (25.03.2018)

[17] Julien Vitay. TU Chemnitz Lecture of Machine Learning, Department of Computer Science, K-Means, PCA, ICA (pp- 5)

[18] Julien Vitay. TU Chemnitz Lecture of Machine Learning, Department of Computer Science, K-Means, PCA, ICA (pp- 15)

[19] Julien Vitay. TU Chemnitz Lecture of Machine Learning, Department of Computer Science, K-Means, PCA, ICA (pp- 6)

[20] Yanyan Meng, University of Nebraska – Lincoln, Digital Commons 4-2012 (pp 56-57)

[21] Aarshay Jain, Analytics Vidhya [Online]. Available: https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/ (26.03.2018)

[22] Tavish Srivastaya, Analytics Vidhya [Online]. Available: https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/ (26.03.2018)

[23] Thomas Vincent, Digital Ocean [Online]. Available: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3 (26.03.2018)

[24] Oruc, Orcun, 2017 "XReview Image Collection". Unpublished image

[25] Aars Thirunav, 2016, Quora Question [Online]. Available: https://www.quora.com/What-is-ARIMA (26.03.2018)

[26] Ecostat Unical (pp-2) [Online]. Available: http://www.ecostat.unical.it/tarsitano/didattica/SeStoCor/SeStor%2027.3/08notes5GOOD.pdf (26.03.2018)

[27] Julian McAuley, Amazon Product Data [Online]. Available: http://jmcauley.ucsd.edu/data/amazon/links.html (26.03.2018)

[28] Julien Vitay. TU Chemnitz Lecture of Machine Learning, Department of Computer Science, K-Means, Recurrent Neural Networks (pp- 9)

[29] Julien Vitay. TU Chemnitz Lecture of Machine Learning, Department of Computer Science, K-Means, Vanishing Gradient Problem (pp- 28)

[30] Liu Sida, Liu Sida's Homepage Recurrent Neural Network [Online]. Available: https://liusida.github.io/2016/11/04/rnn-implementation/ (26.03.2018)

[31] Christopher Olah, colah's Blog [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (26.03.2018)

[32] Adam McCombs, Robert Banh, The Definitive Guide To Magento (pp – 3)

[33] Adam McCombs, Robert Banh, The Definitive Guide To Magento (pp – 5)

[34] Adam McCombs, Robert Banh, The Definitive Guide To Magento (pp – 135, 136)

[35] BlueFoot CMS and Page Builder, Introduction Page [Online]. Available : https://www.bluefootcms.com (27.03.2018)

[36] Oruc, Orcun, 2017 "BlueFoot Image Collection". Unpublished image. (Courtesy of BlueFoot CMS)

[37] Apache Open Source Community, Hadoop [Online]. Available: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html (27.03.2018)

[38] Ahmed Ghandour, Department of MIS, International Journal of Economics, Big Data Driven E-Commerce Architecture (pp – 994)

[39] Jason Brownlee, Machine Learning Mastery Blog [Online]. Available: https://machinelearning-mastery.com/time-series-forecasting-long-short-term-memory-network-python/ (27.03.2018)

[40] Upstream Commerce, Predictive and Dynamic Pricing [Online] Available: http://upstreamcom-merce.com/products_/dynamic-pricing-predictive-analytics/ (27.03.2018)

[41] Blur Admin Template [Online] Available: https://akveo.github.io/blur-admin/ (27.03.2018)

Appendix

```
@app.route('/deepLearning')
def test():
    print("RandomSeedCreator")
    random.seed(100)
    rangeSerie = pd.date_range(start='2006', periods = 150, freq='M')
    print("rangeSerie", rangeSerie)
    ts = pd.Series(np.random.uniform(5, -5, size=len(rangeSerie)),
rangeSerie).cumsum()
    ts.plot(c='b', title='XReview Price Deviation')
    plt.show()
    ts.head(10)

    TS = np.array(ts)
    num_periods = 20
    f_horizon = 1 # forecast horizon

    x_data = TS[:(len(TS) - (len(TS) % num_periods))]
    x_batches = x_data.reshape(-1, 20, 1)

    y_data = TS[1:(len(TS) - (len(TS) % num_periods)) + f_horizon]
    y_batches = y_data.reshape(-1, 20, 1    )
    print(len(x_batches))
    print(x_batches.shape)
    print(x_batches[0:2])

    print(y_batches[0:1])
    print(y_batches.shape)

    def test_data(series, forecast, num_periods):
        test_x_setup = TS[-(num_periods + forecast):]
        testX = test_x_setup[:num_periods].reshape(-1, 20, 1)
        testY =TS[-(num_periods):].reshape(-1, 20, 1)
        return testX, testY

    X_test, Y_test = test_data(TS, f_horizon, num_periods)
    print(X_test.shape)
    print(X_test)

    tf.reset_default_graph()

    num_periods = 20
    inputs = 1
    hidden = 300
    output = 1
```

Figure 2.8 : Sample Deep Learning Test Phase [39]

```
@app.route('/xpred')
def xpred():
    print("API calls")
    random.seed(100)
    rangeSerie = pd.date_range(start='2008', periods = 130, freq='M')
    print("rangeSerie", rangeSerie)
    randomPrices = pd.Series(np.random.uniform(-5, 5, size=len(rangeSerie)),
rangeSerie).cumsum()
    randomPrices.plot(c='b', title='XReview Price Changes')
    randomPrices.head()

    randomPrices = randomPrices.resample('MS').mean()
    randomPrices = randomPrices.fillna(randomPrices.bfill())

    print(randomPrices)

    randomPrices.plot(figsize=(15, 6))
    p = d = q = range(0 , 2)

    pdq = list(itertools.product(p, d, q))

    seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

    warnings.filterwarnings("ignore") # specify to ignore warning messages

    for param in pdq:
        for param_seasonal in seasonal_pdq:
            try:
                mod = sm.tsa.statespace.SARIMAX(randomPrices,
                                                order=param,
                                                seasonal_order=param_seasonal,
                                                enforce_stationarity=False,
                                                enforce_invertibility=False)

                results = mod.fit()

            except:
                continue

    mod = sm.tsa.statespace.SARIMAX(randomPrices,
                                    order=(1, 1, 1),
                                    seasonal_order=(1, 1, 1, 12),
                                    enforce_stationarity=False,
                                    enforce_invertibility=False)

    results = mod.fit()


    pred_dynamic = results.get_prediction(start = pd.to_datetime('2008-01-01'),
dynamic=True, full_results=True)
    pred_dynamic_ci  = pred_dynamic.conf_int()

    y_forecasted = pred_dynamic.predicted_mean
```

Figure 2.9 : ARIMA Process [23]