Mid-project Progress Status Report for Chatbot Project

Client:

Mr. Goldy T Wijaya

Team Members:

AUT ID	Name
20108405	Tianyang Li
20108728	Shijie Ma
20108736	Hanpeng Jiang
20108511	Jiacheng Lv

Mentor Name:

Mr. Petteri Kaskenpalo

Date: 1/3/2022

Version 1.1

Table of Contents

Executive Summary

This is a mid-project progress status report for developing a chatbot for Jaktent conference and virtual exhibition website. The status report begins with an overview of the project's specific objectives, approach, key milestones and deliverables. At the same time, it responds to the client's suggestions for the proposal and lists the new variations from the original proposal. Then, the status report lists the current status of the entire project, divides the unfinished work of the project into the project plan, and puts forward some suggestions for improving the quality of the team's work. The team's individual learning achievements and contributions are also described at the end of the report.

1. Project Description

Chatbots are a popular way to reach out to website users and help them find the answers or information they are looking for. The main goal of this project is to develop an intelligent chatbot for the Jaktent conference website, which can help people to use the website and find out the information they need, and at the same time, the chatbot can quickly solve the problems encountered by people when they browse this website.

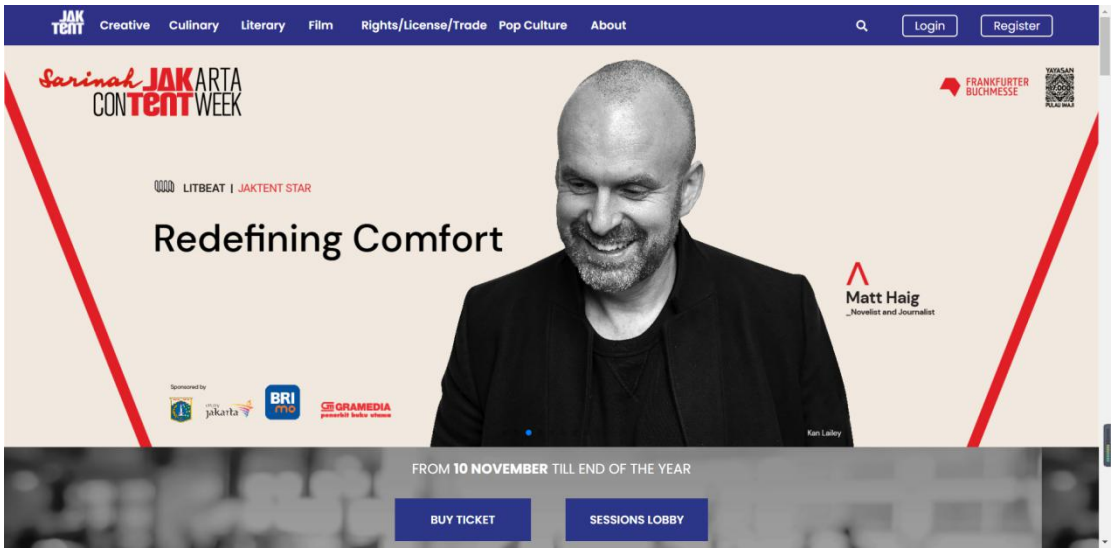


Figure 1: Jaktent Website

There are many types of chatbot frameworks on the market, which use different technologies and have their own advantages and disadvantages. In order to make the whole project more successful, we adopted an incremental research approach, starting with the simplest machine learning models, learning step by step and moving towards advanced deep learning and natural language processing techniques. At the same time, develop several chatbot models using machine learning and deep learning techniques and select the most suitable chatbot model. Then, analyze the content of the Jaktent conference website, crawl certain data as the training corpus of the chatbot, and use the developed chatbot model and front-end technology to finally realize an intelligent chatbot for Jaktent website.

2. Project Scope

2.1. Objectives

The goal of the project is to create a chatbot for a virtual conference and exhibition called Jaktent, and use its conference materials and contents as training and examples for such chatbot configuration. This chatbot can help people use the website and find out useful information they need.

2.2. Approach

The chatbot project is developed using agile development methods, which are in line with the gradual research and experimental development model of the entire project. At the same time, it can adapt to the changing needs of customers, and can quickly find the loopholes in the project and deviate from the project goals.

In addition to the overall agile development model of the project, we adopted an incremental development research method throughout the project development process. Start with simpler machine learning techniques and step by step to explore more advanced chatbot frameworks based on deep learning and natural language processing. After developing several chatbot models, we will choose the most suitable model, cooperate with crawler and front-end technology, and finally develop a more intelligent chatbot for Jaktent website.

2.3. Major Milestones

According to the project plan, in the nearly 8-month project cycle, the entire project is mainly divided into 5 stages, and the milestones corresponding to each stage are shown in the following table:

Project Stage	Milestone
Project Kick-off	Complete the project proposal and give a project proposal presentation in early November
Project Phase 1: Basic Knowledge Learning + Chatbot Threshold	(1) Basically master the relevant computer technology related to the development of chatbot, and show

	learning results through case study. (2) Master and test some simple chatbot frameworks (mainly based on machine learning)
Project Phase 2 -- Different Chatbot Framework Testing and Experimental Prototype Development	(1) Develop several chatbot prototypes and test their performance (2) Select the best chatbot model after designing several chatbot prototypes
Project Phase 3 -- Jaktent Conference Website Chatbot Design and Development	Complete the development of chatbot on Jaktent conference website
Project Phase 4 -- Project Test and Final Presentation	Completely show the finished chatbot to the client, and finally hand in all the required assessment documents

2.4. Deliverables

(1) Project management-related deliverables: scope statement, project schedule, status reports, final project presentation, final project report, individual reflective report, and any other documents required to manage the project.

(2) Product-related deliverables: project prototype, design documents, project instruction, source code, etc.

3. Response to Proposal Recommendations

(1) We need to identify the specific technologies we need to use for our project in the proposal.

Response: Through the client meeting after the proposal and combined with our own actual situation, we finally determined the development technology of our whole project. The front end uses web development language (HTML + CSS + JavaScript) and for the back end, considering that Python has great advantages in data processing and data analysis, we finally decided to use Python as the main development language

and combined with machine learning, deep learning and natural language processing technology to develop intelligent chatbot.

(2) The specific work allocation of the project needs to be added to the proposal.

Response: Considering that our project is generally a research project, which is different from engineering projects, at first we only specified one team member to be responsible for the front-end and subsequent front-end and back-end interaction, and the other three team members are responsible for the back-end chatbot model building, data mining, corpus building, etc. Due to the great correlation between Python data analysis and various parts of machine learning and we lack some foundation of machine learning and deep learning, at first, we spent some time learning the basic knowledge of machine learning, deep learning and natural language processing. After we have a certain knowledge and in-depth understanding, we allocate the project work, as shown in the table below:

Team Member	Responsible Work
Tianyang Li	Team leader, mainly responsible for project work allocation, project process coordination and coordinating client meeting time with client and mentor. Machine learning and NLP Developer, mainly responsible for the development of chatbot model and relevant data processing.
Hanpeng Jiang	Machine Learning and NLP Developer, mainly responsible for the development of chatbot model and relevant data processing.
Shijie Ma	Data Mining and Analysis Developer, mainly responsible for data crawling, data cleaning and corpus construction of Jaktent conference website, and coordinating the front-end and back-end interaction of the project.
Jiacheng Lv	Front-end Developer, mainly responsible for developing the front-end interface of the chatbot and coordinating the front-end and back-end interaction of the project.

4. New Variation from the Proposal

There is no problem with the overall project plan and schedule, and the project is running smoothly as a whole. Some major new variations are as follows:

(1) Crawl the data of Jaktent conference website through crawler technology, and form a specific corpus of chatbot through data cleaning.

At first, we thought that the website has a ready-made corpus that can be provided to us as a training corpus for chatbot development. However, through communication with customers, because Jaktent website involves a wide range of contents, client needs us to analyze the website and obtain valuable information and data for chatbot development. Therefore, we are scheduled to use crawler technology to crawl the relevant data of the website, and build a unique corpus of the website through data cleaning.

(2) Modify the existing chatbot front-end framework according to the actual situation, thereby reducing the time spent on front-end development of the project.

In our initial plan, we needed to develop a front-end interface for our chatbot, for which we also designed a prototype diagram, but the advice given to us by our mentor is to study some open source chatbot frameworks and make modifications on this basis, and finally form our own front-end interface, which can reduce the time required for front-end development of the project. In this regard, we are prepared to work on the advice of our mentor.

(3) Combine with some open source chatbot frameworks to further make the chatbot more intelligent.

In the previous stage, we have developed several chatbot models using machine learning, deep learning and natural language processing technologies, and have basically realized the daily chat function through corpus training. In the research, we found that the effect of chatbots will be better when the corpus is sufficient, but some professional problems still require some professional corpus. These results have laid a good foundation for us to finally complete the Jaktent conference website chatbot development. At the same time, the mentor suggested that we can combine some open

source chatbot frameworks (e.g. Botonic, DeepPavlov etc.) to make our chatbot more intelligent. In the next stage, while the project is developing, we will also make some attempts in this regard at the same time.

(4) Use a combination of database indexing and chat to build the ultimate Jaktent conference website chatbot.

At first we had a lot of doubts about the type of chatbots for the Jaktent conference website, because if the corpus is mixed with website content and daily chat, it will increase the difficulty of chatbot training. By communicating with the mentor, the mentor suggested that we can divide the corpus into two parts, one is the corpus database of website content, and the other is the daily chat words, which can make the training of the subsequent model easier.

5. Project Status

5.1. Completed Work

Completed Work	Description
Proposal Writing and Presentation	Through thesis reading, research on the project background, and through the client meeting with the client to understand the client's requirements. After research, identify the project's objectives, development methods, technical requirements, make the project plan, write the proposal and make the presentation.
Basic Knowledge Learning	Learn the basic knowledge relevant to the chatbot project, including: (1) Basic knowledge of web development (HTML+CSS+Javascript) (2) Python review (3) Machine learning and deep learning (4) Natural language processing and data mining
Research on Chatbot	Conduct research and some learning tests on common

Model and Testing	chatbot models, including: (1) Machine learning based chatbot model (e.g. Decision Tree, Bayesian Classifier, SVM, etc.) (2) Deep learning based chatbot model (e.g. Tensorflow, RNN, Seq2Seq, etc.) (3) (1) (2) are mainly based on retrieval-based chatbots, and a series of studies on generative chatbot models have also been conducted Combined with the actual situation of the Jaktent conference website, we believe that it is more practical to make a retrieval-based chatbot. The chatbot trained through a certain Jaktent website corpus can solve the problems encountered by customers when using the website.
Machine Learning Based Chatbot Model 1	A total of two models have been developed for chatbots based on machine learning. The first one is based on decision trees.
Machine Learning Based Chatbot Model 2	Based on Bayesian Classifier
Deep Learning Based Chatbot Model 1	The chatbot based on deep learning has also developed a total of two models, the first one is based on Tensorflow+RNN+LSTM
Deep Learning Based Chatbot Model 2	Based on Tensorflow+Seq2Seq (A variant of RNN)

5.2. Uncompleted Work

Uncompleted Work	Description	Estimated Completed Time
Front-end Interface	The development of the front-end interface	2022.3.20

Development	of the chatbot is already in progress. According to the mentor's suggestion, we will study some open-source chatbot front-end frameworks and make modifications on this basis to form our own front-end interface, thus saving the time spent in this part.	
Corpus Building	This part has also started, mainly using crawling technology to crawl the content of the Jaktent conference website, and forming a corpus for training our chatbot model through data mining and data cleaning technology. Overall, this part is very critical, which will directly affect subsequent model training and testing.	2022.3.27
Enhance the Performance of Chatbot Model	Combined with some open-source chatbot frameworks, the performance of the chatbot model is further improved, and the cleaned corpus is used for training and testing.	2022.4.7
Subtasks combine to build the final product	Combine a series of modules to build a complete Jaktent conference website chatbot with front-end interactive interface, and test and improve it according to the actual situation.	2022.5.8
All project related document preparation, defense preparation and final submission	All project-related documents are submitted by the end of May and the project is finalized.	2022.5.31

According to the overall progress of the project, after the mid-project progress

review, the whole project will enter in the fourth stage. We further subdivide the project plan for the next step, as shown in the figure below:

任务模式	任务名称	工期	开始时间	完成时间	前置任务	基线完成时间
	4.0 Project Phase 3 -- Jaktent Conference Website Chatbot Design and Development	**34 个工作日**	**2022年3月5日**	**2022年4月20日**	**17**	**2022年4月20日**
	4.1 Chatbot Front-end Interface Development	12 个工作日	2022年3月5日	2022年3月20日		2022年3月20日
	4.2 Building Jaktent Website Corpus--Data Crawling, Mining and Cleaning	16 个工作日	2022年3月7日	2022年3月27日		2022年3月27日
	4.3 Enhance the Performance of the Chatbot Model	18 个工作日	2022年3月7日	2022年3月30日		2022年3月30日
	4.4 Model Testing with Jaktent Website Corpus	5 个工作日	2022年4月1日	2022年4月7日	27,28	2022年4月7日
	4.5 Combine All the Subtask to Build the Initial Product	9 个工作日	2022年4月10日	2022年4月20日		2022年4月20日
	5.0 Project Phase 4 -- Project Test and Final Presentation	**29 个工作日**	**2022年4月21日**	**2022年5月31日**	**25**	**2022年5月31日**
	5.1 Project Testing and Improvement	13 个工作日	2022年4月21日	2022年5月8日		2022年5月8日
	5.2 Final Poster and Portfolio Preparation	11 个工作日	2022年5月9日	2022年5月22日	32	2022年5月22日
	5.3 Final Presentation	1 个工作日	2022年5月25日	2022年5月25日	33	2022年5月25日
	5.4 Individual Reflective Report Writing	4 个工作日	2022年5月26日	2022年5月31日	34	2022年5月31日

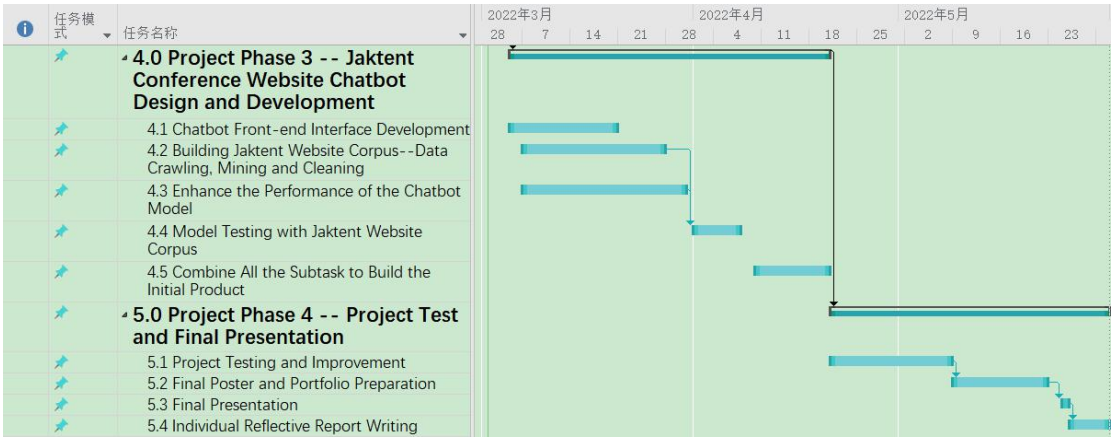


Figure 2: Project Plan for the Next Step

6. Existing Issues in Project and Team

Issue Description	Impact on Project	Comments
Have trouble with the environment establishment of Python	The code couldn't run	Learn more details about establishing environment
Wrong version of CUDA	Tensorflow cannot be used	Deeper understanding of the compiling environment
The version of Tensorflow was too advanced	Some model from internet cannot operate normally	Learned more about Tensorflow 2.0
The customer could not	Unable to implement	Learn python crawler

provide a corresponding corpus for the chatbot	efficient database retrieval	technology to crawl web content
Unclear division of roles	Group can't be productive	Have team meetings for a clearer division of labor

7. Recommendations for Improving Team's Performance

(1) Strengthen communication with client

Sort out the current status of the project and communicate with client. At present, the team has improved in this aspect. After that, we will continue to communicate with client in the important stage of the project in time, listen to client's opinions and suggestions, and ensure that the project meets client's needs.

(2) Regular meetings within the team

Team meetings need to be more frequent and regular. During the meeting, project members report the current progress and later plan. In addition, strive to strengthen team cooperation and improve work efficiency.

(3) Try to use Github to manage the team project

After finishing the current work, upload the work file to GitHub and inform team members of any significant progress or problems.

(4) Enhance teamwork and communication on each person's different projects

Although everyone is responsible for different sectors, each sector is interconnected. For example, front-end programmers should establish contact with back-end programmers to solve some system problems together. So teamwork is very important in our development process.

8. Individual Contributions & Learning Achieved

Tianyang Li

As the team leader of the entire project, my main task is to complete the allocation of the daily project development work of the team, coordinate the work process of the entire project, and discuss and modify the specific project plan with

team members when necessary, so as to conform to the characteristics of agile development. At the same time, I am also responsible for daily communication with clients and mentors, as well as organizing daily client meetings and coordinating team member and client time.

During the project development process, I was mainly responsible for the development of the chatbot model and part of the data mining and processing. Before the start of the project, I didn't have much basic knowledge in machine learning, so I began to learn from the most basic machine learning knowledge, learned some basic machine learning algorithms, and developed two chatbot models based on machine learning with my team using Decision Tree and Bayesian classification algorithm. At the same time, we found that this predefined chatbot model can only carry out matching answers according to the corpus, but can not learn through the corpus. At the same time, it can only use small corpus, and the effect is very poor. Therefore, after that, I began to learn some knowledge about deep learning and natural language processing, and finally developed a relatively intelligent chatbot model by using Tensorflow + Seq2Seq. The chatbot model can learn through a large number of corpus, and can basically carry out daily chat after training.

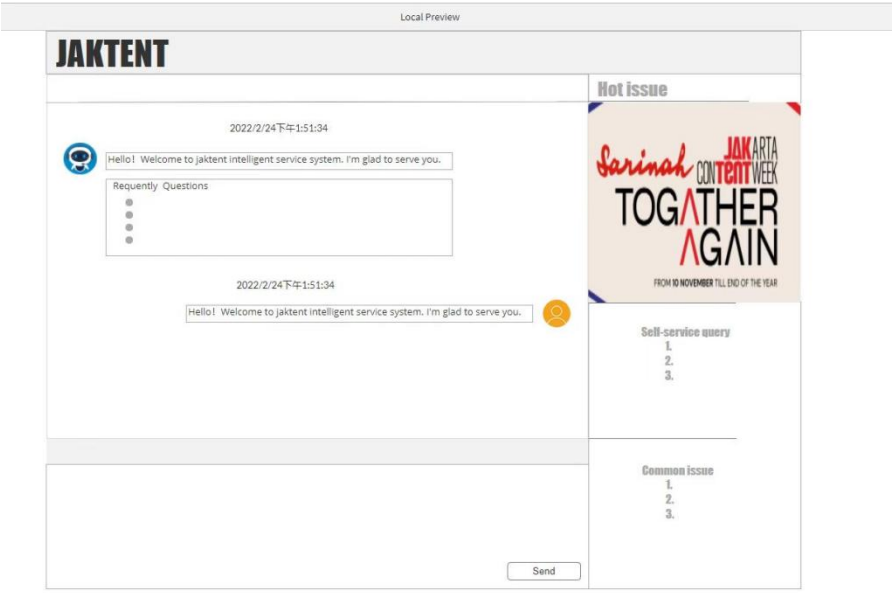
Hanpeng Jiang

During the first half of R&D project, I was mainly responsible for building a chatbot model that can reply to the questions automatically. I read some papers on chatbots and increased my understanding of how chatbots work. I learned Python, machine learning, deep learning(CNN, RNN, LSTM) to have a good comprehension of the whole project. I tried some machine learning algorithms to increase my awareness of chatbot models. Also, I chose RNN+Tensorflow to develop one of our chatbot models. The model is also trained with some algorithms like LSTM to make the chatbot's response more accurate and intelligent.

Shijie Ma

In this R&D project, I was mainly responsible for the building of the prototype of the web page and the part of web page art in the early stage. I used Axure to build a preliminary prototype of the web page, and built the basic prototype of the page of the

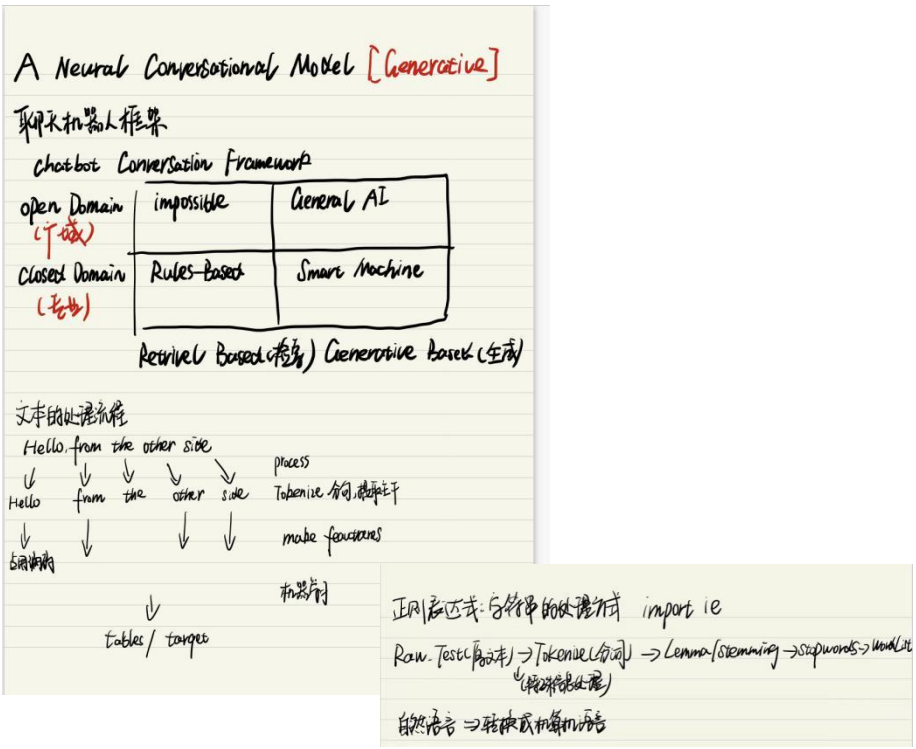
intelligent chatbot. In the later stage, I used the ready-made web chat framework to modify it into a suitable template for our web chat.



In addition, I also learned the process and basic principles of chatbot construction using deep learning and machine learning. At the same time, since the relevant person in charge of Jaktent website could not provide the database of relevant webpage content, I also need to use scrapy crawler framework to extract and classify the website content.

Jiacheng Lv

I was in charge of the Web development task and coordinating the front-end and back-end interaction of the R&D project. Due to my limited knowledge about Web, I spent more than a month to systematically learn HTML, CSS, Javascript and Vue framework in Web development, from only building static Web pages to learning how to establish connections with back-end databases. I learned a lot from it. In the current R&D project, I have built an initial framework of intelligent chatbot. In the later stage, I will beautify the web page with CSS and add some dynamic effects to the web page with JS, so that the web page of intelligent chatbot can be completely developed. Then I communicated with the students in charge of the back-end and learned how to connect the front-end Web page with the back-end database. In this process, I gained a lot from them.



tf-idf

```
In [44]: from sklearn.feature_extraction.text import TfidfVectorizer
         vectorizer = TfidfVectorizer(min_df=10,
                                      max_features=180000,
                                      tokenizer=tokenize,
                                      ngram_range=(1, 2))

In [45]: all_desc = np.append(train['item_description'].values, test['item_description'].values)
         vz = vectorizer.fit_transform(list(all_desc))

In [46]: vz.shape
Out[46]: (2175890, 180000)

In [47]: #  create a dictionary mapping the tokens to their tfidf values
         tfidf = dict(zip(vectorizer.get_feature_names(), vectorizer.idf_))
         tfidf = pd.DataFrame(columns=['tfidf']).from_dict(
                             dict(tfidf), orient='index')
         tfidf.columns = ['tfidf']
```

2021.12.6--12.19

In the past two weeks, we focused on the homework and assessment of "network and system administration" and "information security management", so the follow-up of the project was temporarily suspended for two weeks.

2021.12.21

Have a preliminary understanding of the development history of artificial intelligence, neural network and the related framework and application of machine learning and deep learning, and learn how to install tensorflow 2.0 on the computer.

2021.12.22

Learned an important problem of machine learning - regression problem, including linear regression, gradient descent and logical regression, and practiced the gradient descent and update algorithm through a practical example of linear model. The following is a part of the code I wrote during learning.

```
def step_gradient(b_current, w_current, points, learningRate):
    b_gradient = 0
    w_gradient = 0
    N = float(len(points))
    for i in range(0, len(points)):
        x = points[i, 0]
        y = points[i, 1]
        # grad_b = 2(wx+b-y)
        b_gradient += (2 / N) * ((w_current * x + b_current) - y)
        # grad_w = 2(wx+b-y)*x
        w_gradient += (2 / N) * x * ((w_current * x + b_current) - y)
    # update w'
    new_b = b_current - (learningRate * b_gradient)
    new_w = w_current - (learningRate * w_gradient)
    return [new_b, new_w]


def gradient_descent_runner(points, starting_b, starting_w, learning_rate, num_iterations):
    b = starting_b
    w = starting_w
    # update for several times
    for i in range(num_iterations):
        b, w = step_gradient(b, w, np.array(points), learning_rate)
    return [b, w]
```

Figure 3: Individual Logbook Example