**PRESENTATION SCRIPT**

# **I/ TABLE OF CONTENT**

[**I/ TABLE OF CONTENT 1**](#_heading=h.gjdgxs)

[**II/ TABLE OF FIGURES 1**](#_heading=h.30j0zll)

[**III/ LIST OF TABLES 2**](#_heading=h.1fob9te)

[**IV/ INTRODUCTION 2**](#_heading=h.3znysh7)

[**V/ PRESENTATION SCRIPT 2**](#_heading=h.2et92p0)

[**1/** About us](#_heading=h.3dy6vkm) 3

[**2/** Problem – solving & Idea](#_heading=h.17dp8vu) 5

[**3/** Technical Craftsmanship](#_heading=h.44sinio) 8

[**4/** UI – UX](#_heading=h.ihv636) 19

[**5/** Product future](#_heading=h.vx1227) 22

[**6/** Struggle](#_heading=h.4f1mdlm) 23

[**VI/ REFERENCES 25**](#_heading=h.28h4qwu)

# **II/ TABLE OF FIGURES**

[Figure 1: Table of Content in slide 3](#_heading=h.tyjcwt)

[Figure 2: About us 3](#_heading=h.1t3h5sf)

[Figure 3: µεράκι members and roles 4](#_heading=h.4d34og8)

[Figure 4: What sets us apart? 5](#_heading=h.2s8eyo1)

[Figure 5: Problem-solving and Idea 5](#_heading=h.3rdcrjn)

[Figure 6: Problem - solving of FSI & EPU 6](#_heading=h.26in1rg)

[Figure 7: Idea of problem 6](#_heading=h.lnxbz9)

[Figure 8: Usecase diagram 7](#_heading=h.35nkun2)

[Figure 9: Flowchart 8](#_heading=h.1ksv4uv)

[Figure 10: Technical craftsmanship 9](#_heading=h.2jxsxqh)

[Figure 11: Technical craftsmanship: Application & Machine Learning 9](#_heading=h.z337ya)

[Figure 12: Application - Frontend 10](#_heading=h.3j2qqm3)

[Figure 13: Frontend – Interface (Login, create campaign) 11](#_heading=h.1y810tw)

[Figure 14: Frontend - Interface (Total numbers of posts, comments, likes) 11](#_heading=h.4i7ojhp)

[Figure 15: Frontend - Interface (Top 5 positive/ negative comments) 12](#_heading=h.2xcytpi)

[Figure 16: Frontend - Interface (The total of comments - relabel) 13](#_heading=h.1ci93xb)

[Figure 17: Application - Backend - Database 13](#_heading=h.3whwml4)

[Figure 18: Steps in Machine learning 14](#_heading=h.2bn6wsx)

[Figure 19: Step 1: Pre - processing (Data cleaning) 14](#_heading=h.qsh70q)

[Figure 20: Processing methods: BoW & Tf - idf 15](#_heading=h.3as4poj)

[Figure 21: Step 2: Processing – Architecture Word2vec 16](#_heading=h.49x2ik5)

[Figure 22: Step 2: Processing - Result of Word2vec 17](#_heading=h.2p2csry)

[Figure 23: Step 3: Build Model - RNN & LSTM 17](#_heading=h.147n2zr)

[Figure 24: Step 3: Build model - Architecture 18](#_heading=h.3o7alnk)

[Figure 25: Step 4: Fine-tuning and final result 19](#_heading=h.23ckvvd)

[Figure 26: UI - UX 20](#_heading=h.32hioqz)

[Figure 27: UI - UX - Login and Create a campaign 20](#_heading=h.1hmsyys)

[Figure 28: UI - UX - The total number of posts, comments, likes – Statistic of](#_heading=h.41mghml)

[positive, negative and neutral comments - Top 5 negative and positive comment 21](#_heading=h.41mghml)

[Figure 29: UI - UX - Comment summary table and re-label 21](#_heading=h.2grqrue)

[Figure 30: Product future 22](#_heading=h.3fwokq0)

[Figure 31: Product future - Model and App 22](#_heading=h.1v1yuxt)

[Figure 32: Struggle 23](#_heading=h.2u6wntf)

[Figure 33: Struggle - Technical & Communication 24](#_heading=h.19c6y18)

[Figure 34: Thanks for listening 25](#_heading=h.3tbugp1)

# **III/ LIST OF TABLES**

[Table 1: Shortcomings of BoW & Tf - idf 16](#_heading=h.1pxezwc)

# **IV/ INTRODUCTION**

Team: µεράκι team

Member:

* Nguyễn Thị Thanh Hường – Product management
* Nguyễn Ngọc Trường Sơn – React Native
* Nguyễn Đức Minh – React Native
* Nguyễn Ngọc Bảo Trân – Data Science
* Đỗ Nam – Data Science

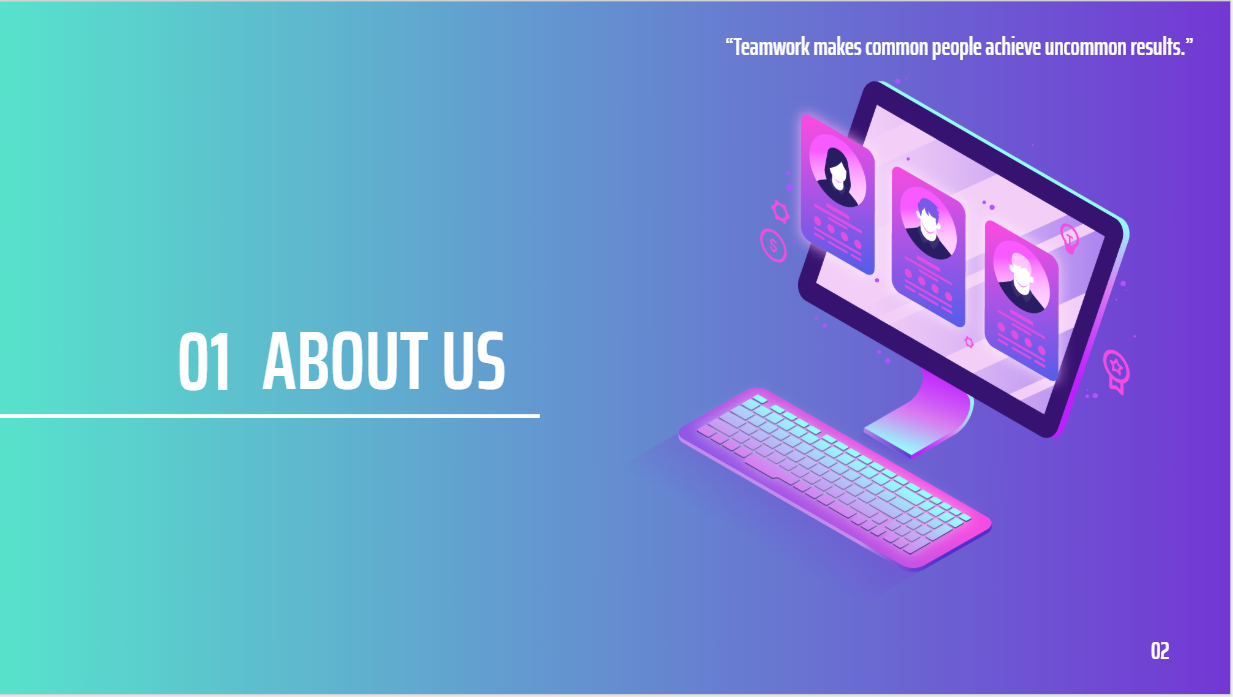
# **V/ PRESENTATION SCRIPT**

The presentation will include 6 main parts. In the first part, I will introduce the team members, each person's role, and the team's purpose in joining the project. The second part is the topic our team chooses to solve and our ideas for the problem. The third part is the technical craftsmanship like frontend, backend, and data that the team will use to process the problem. The next is the UI - UX section. Part 5 is future products: New features and improvements in the future. Last but not least, it's struggle section.



*Figure 1: Table of Content in slide*

## **1/ About us**

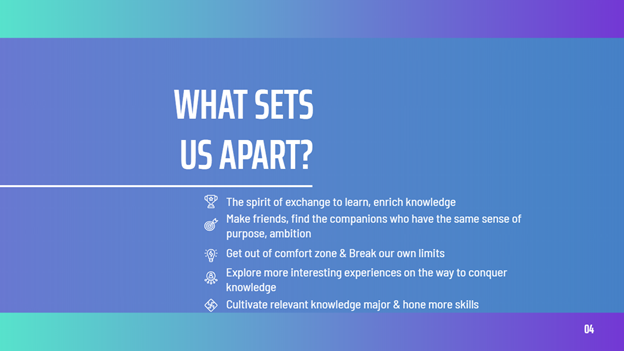


*Figure 2: About us*



*Figure 3: µεράκι members and roles*

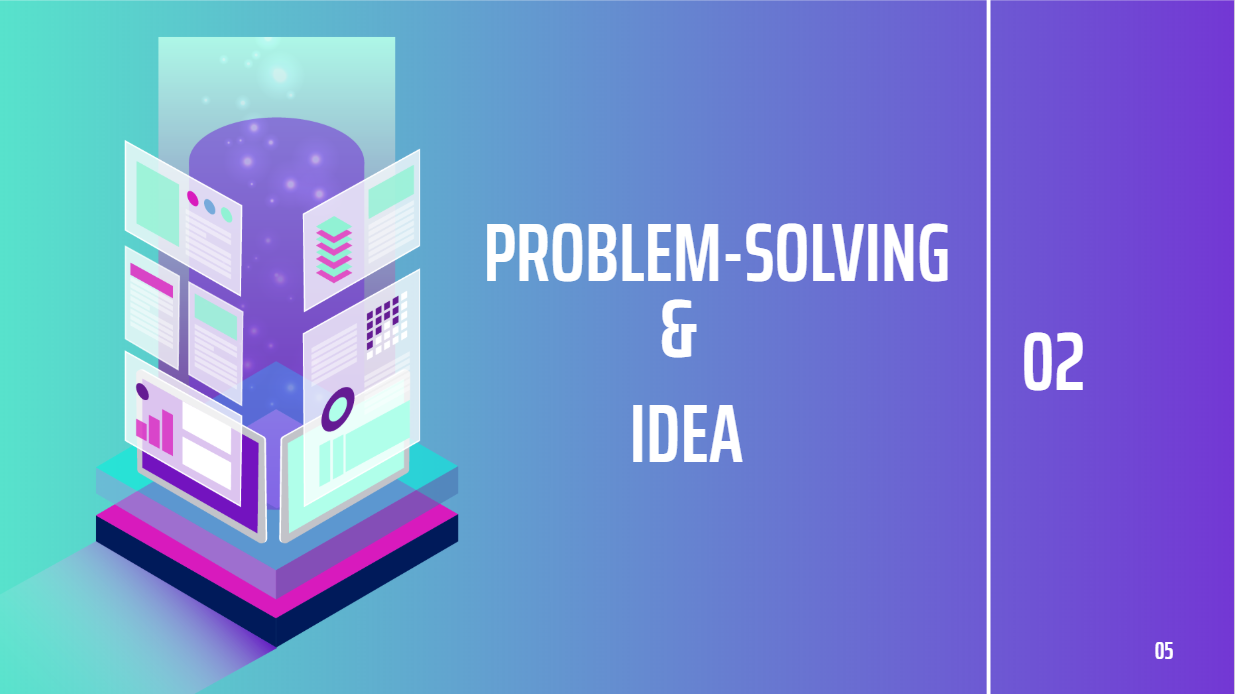
Firstly, I will introduce my team: µεράκι team - there are 5 members including me - Nguyen Thi Thanh Huong - Project manager, 2 data science members Nguyen Ngoc Bao Tran and Do Nam, followed by 2 members react native is Nguyen Duc Minh and Nguyen Ngoc Truong Son. µεράκι comes to DevC with the motto "Alone we can do so little, together we can do so much".



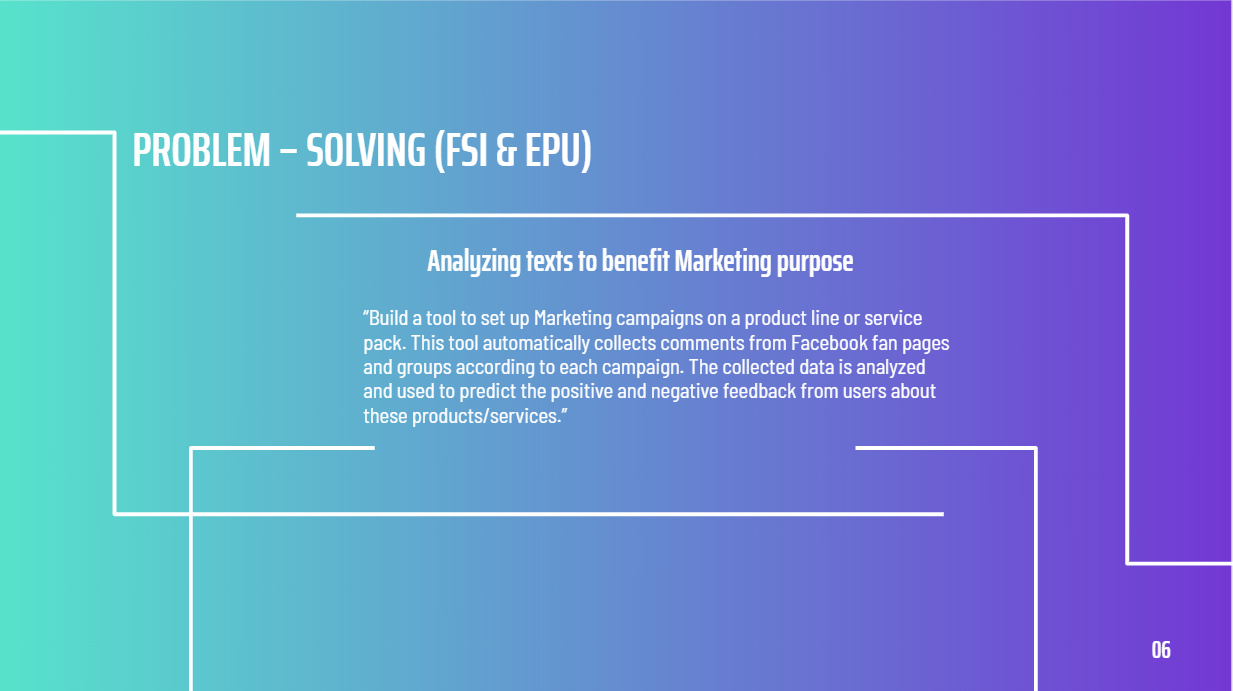
*Figure 4: What sets us apart?*

We come here with the spirit of exchange, learning, enriching our own knowledge. Looking forward to finding more friends, companions with the same purpose, and ideal. Have the opportunity to get out of the inherent comfort zone, to break your limits. Above all, discover more interesting experiences on the way to conquer knowledge, hone practical skills, and cultivate more relevant knowledge of the profession you pursue.

## **2/ Problem – solving & Idea**

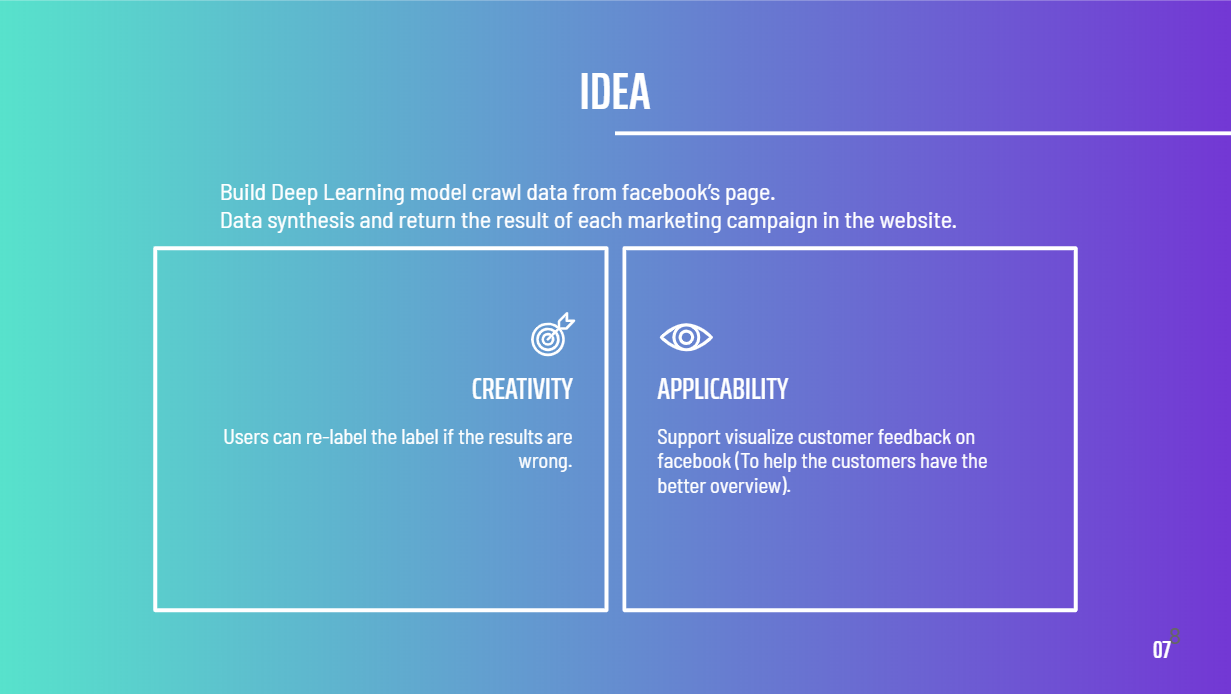


*Figure 5: Problem-solving and Idea*



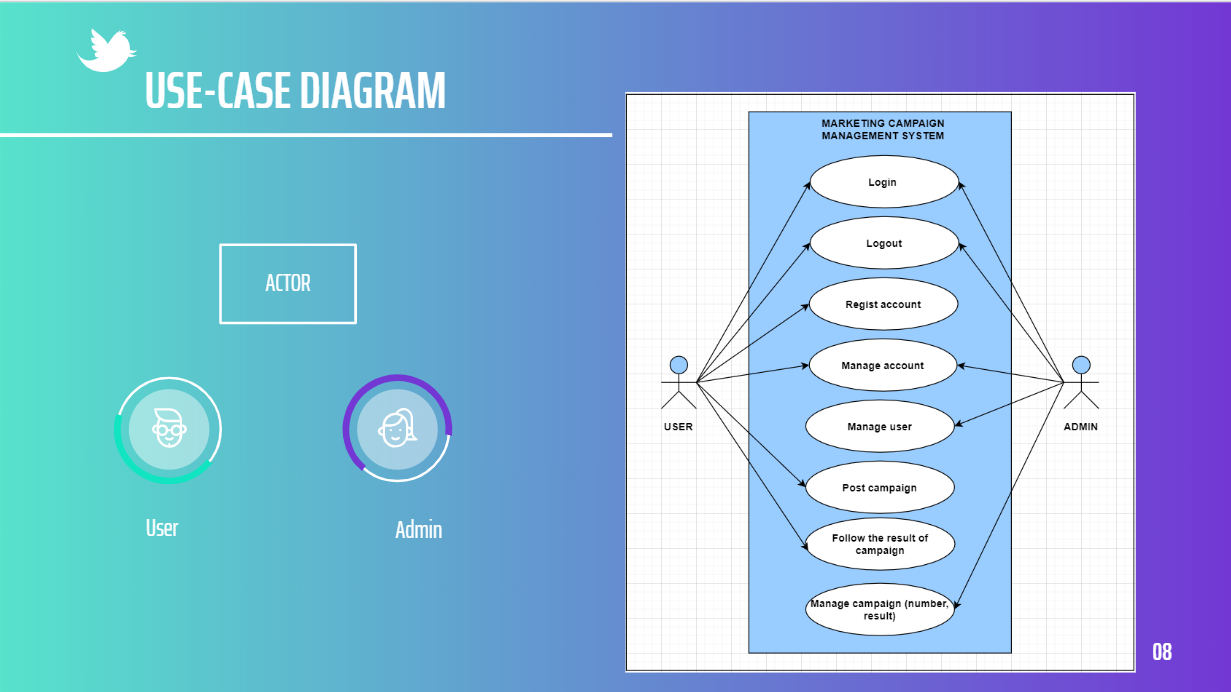
*Figure 6: Problem - solving of FSI & EPU*

In terms of problem-solving and idea, It was very excited and chose the FSI & EPU company's topic "Analyzing documents for marketing purposes".



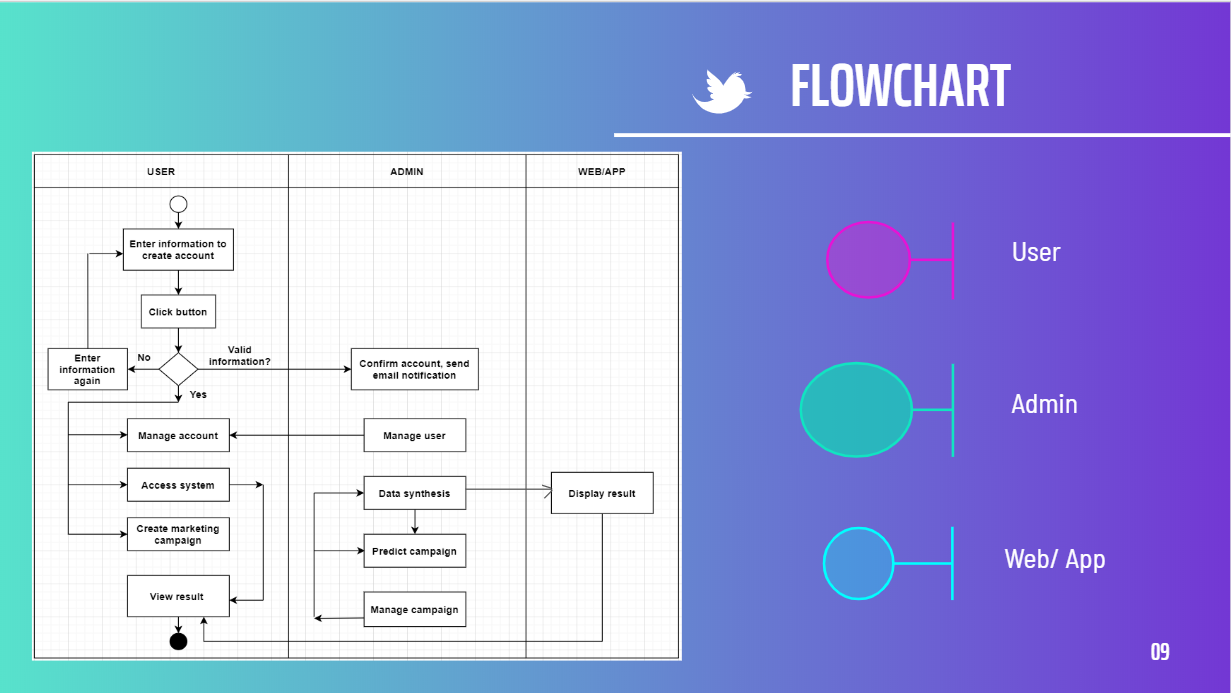
*Figure 7: Idea of problem*

With this problem, we will build a Deep Learning model to crawl data from Facebook to collect data and return the results of each campaign on the website. In idea part, our creativity is shown in the section where the user can relabel the label (display all comments and results) if the label is wrong. In terms of applicability, visual feedback of customers on Facebook can be supported so that customers can have an overview and a holistic view.



*Figure 8: Usecase diagram*

This is the Usecase diagram of the problem, the simulation use-case will have two main objects: User and admin. Both users and admin have the right to login into the system, log out from the system, and manage their accounts. In terms of functions of users who can also register accounts, post campaigns, track results. The admin also has the function to manage users, manage campaigns (quantity, returned results).



*Figure 9: Flowchart*

In addition, It is the flowchart of the system. The first step, after accessing, user will enter information to create an account, click the button. Here, in case the user enters incorrectly, the system will request to re-enter the information to register the account. Information is entered correctly, the account will be created, the system will confirm the account and notify the user by email. User accounts will be managed by the admin. After having an account, users can access the system, manage the account, and create campaigns. After the campaigns are created, the system will manage, predict, and collect data. The collected data will be displayed as a result on the website for the user to see.

## **3/ Technical Craftsmanship**

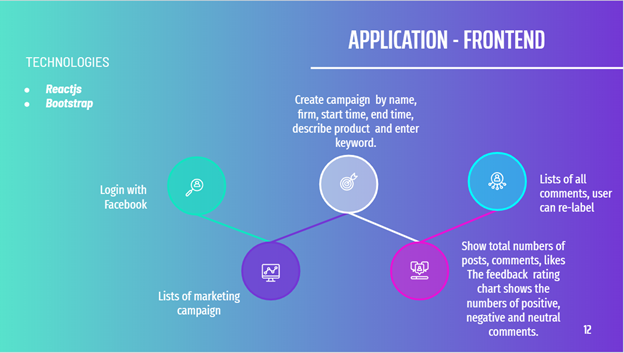


*Figure 10: Technical craftsmanship*



*Figure 11: Technical craftsmanship: Application & Machine Learning*

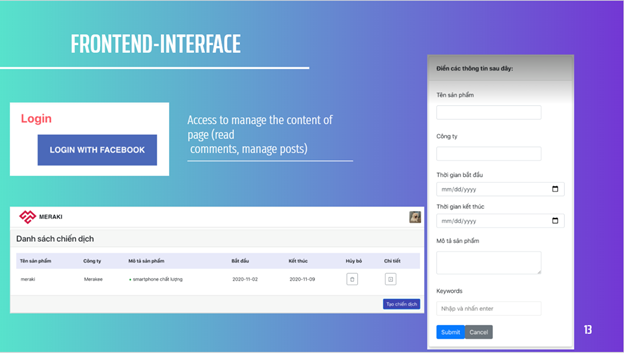
Technology includes application and machine learning. Applications consist of frontend, backend, and database. Machine learning will build a deep learning model.



*Figure 12: Application - Frontend*

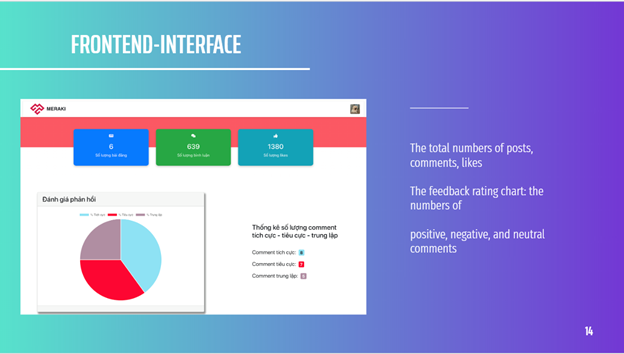
The frontend, the technology chosen by the team is reactjs, the framework that uses bootstrap.

Users will log in with Facebook, create campaigns in the form of name, company, start time, end time, product description, and enter keywords. The interface will also display a list of all user-created campaigns, the total number of posts, comments, and likes. Moreover, there is also a statistical chart of the number of positive, negative, and neutral comments. And one of the special parts will display all the comments and the user can relabel if the label is wrong.



*Figure 13: Frontend – Interface (Login, create campaign)*

Interface: Login with Facebook, create campaign form.



*Figure 14: Frontend - Interface (Total numbers of posts, comments, likes)*

The interface displays the number of posts, comments, likes and graphs, statistics of the number of positive, negative, and neutral comments.



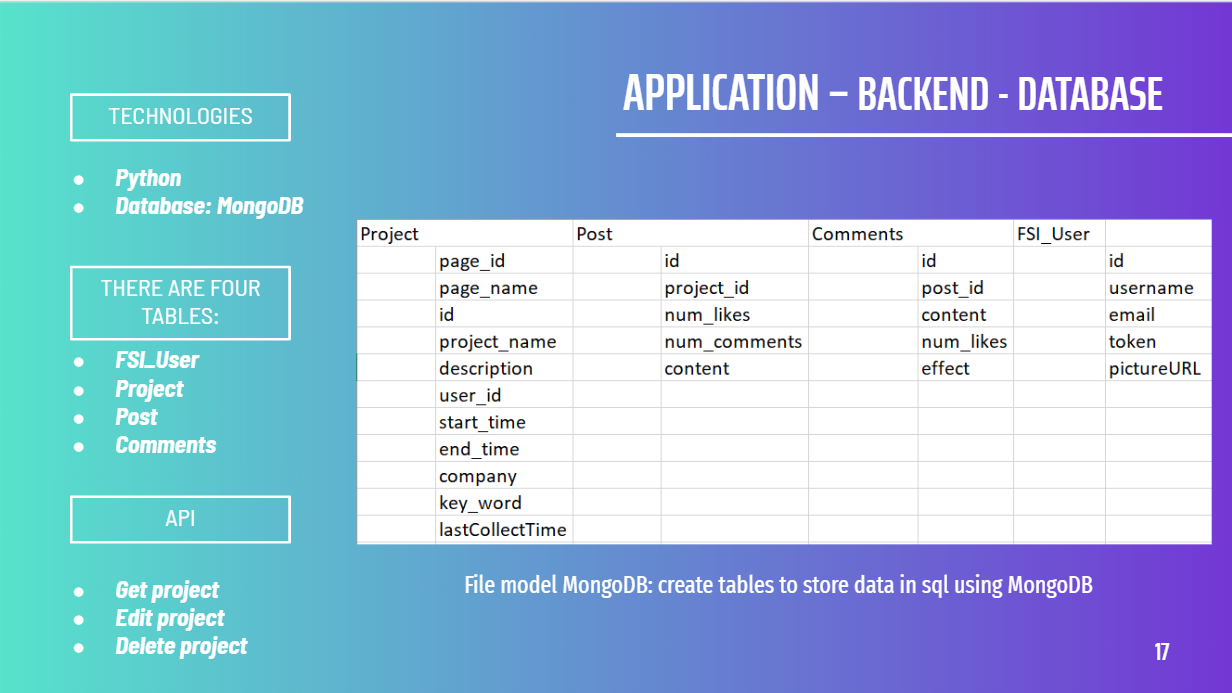
*Figure 15: Frontend - Interface (Top 5 positive/ negative comments)*

Top 5 positive comments and top 5 negative comments.

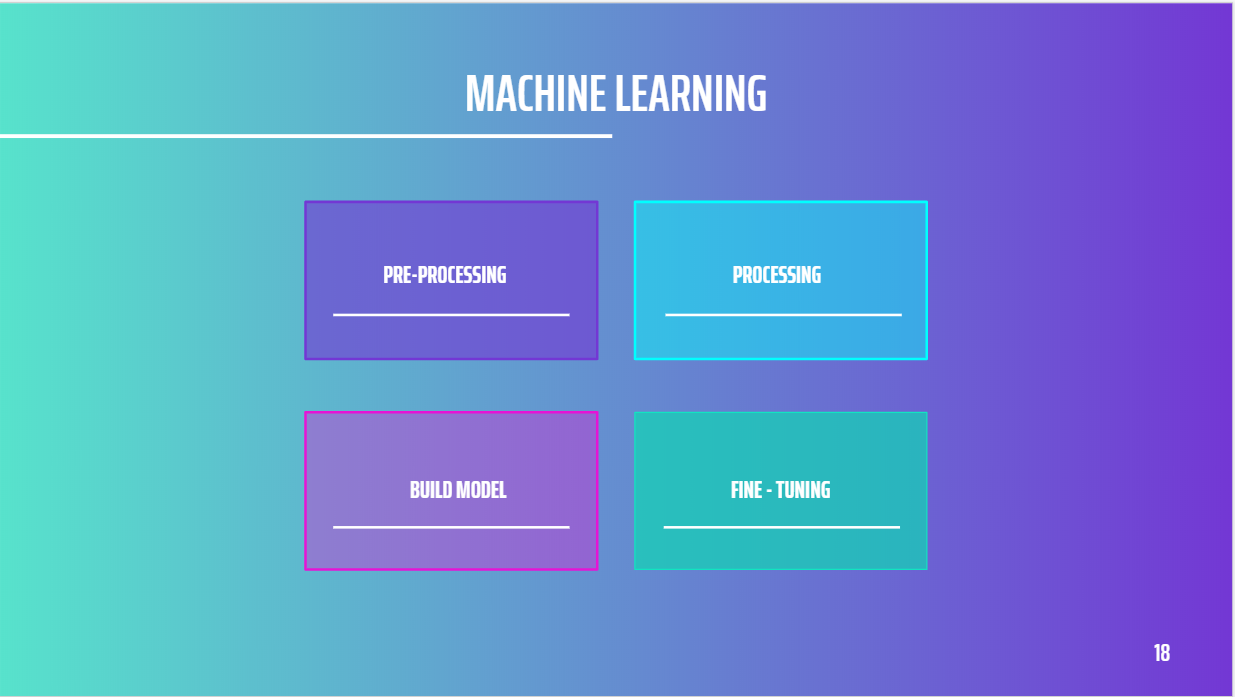


*Figure 16: Frontend - Interface (The total of comments - relabel)*

Summary table of comments, users can update, relabel if the label is wrong.



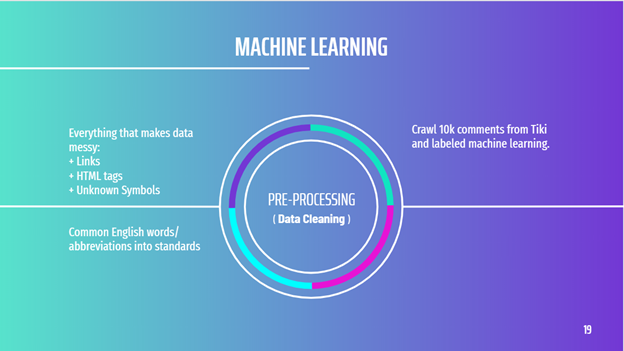
*Figure 17: Application - Backend - Database*

The backend technology used is the python, database: MongoDB, creating a table to store data in SQL using MongoDB with 4 tables: FSI\_User, Project, Post, Comments. 

*Figure 18: Steps in Machine learning*

The next is sentiment analysing, include 4 steps:

* Step 1: Pre – processing
* Step 2: Processing
* Step 3: Build model
* Step 4: Fine – tunning



*Figure 19: Step 1: Pre - processing (Data cleaning)*

Step 1: Pre - processing or data cleaning, the main purpose of this step is removing noise in data such as links, html tags, etc and tokenize words.



*Figure 20: Processing methods: BoW & Tf - idf*

Step 2 is Processing, we'd tried 2 processing methods, BoW and Tf - idf. However, both of these methods still have limitations:

|  |  |
| --- | --- |
| BoW | Tf - tdf |
| Ignore the meaning of each word and  the order of them in the whole sentence.  Large storage vector. (Equal to the size  of vocabulary set.) | Even though the TF-IDF shows the weights of words in the text, this method still does not represent order of words in sentence and still keeps the large feature vector as learning material. |

*Table 1: Shortcomings of BoW & Tf - idf*

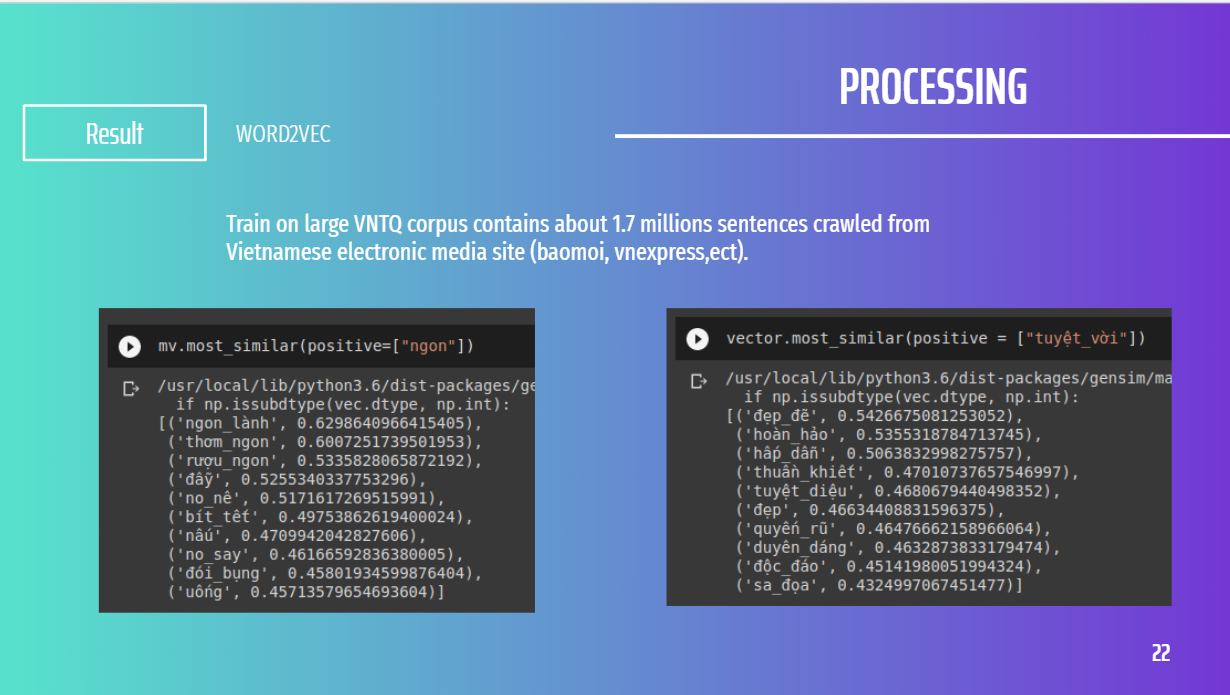


*Figure 21: Step 2: Processing – Architecture Word2vec*

As a result, the accuracy of these methods are not so convincing. So we have 2 problems here are: The order-of-words problem and the storage-vector problem. To solve these, we tried word2vec and LSTM.

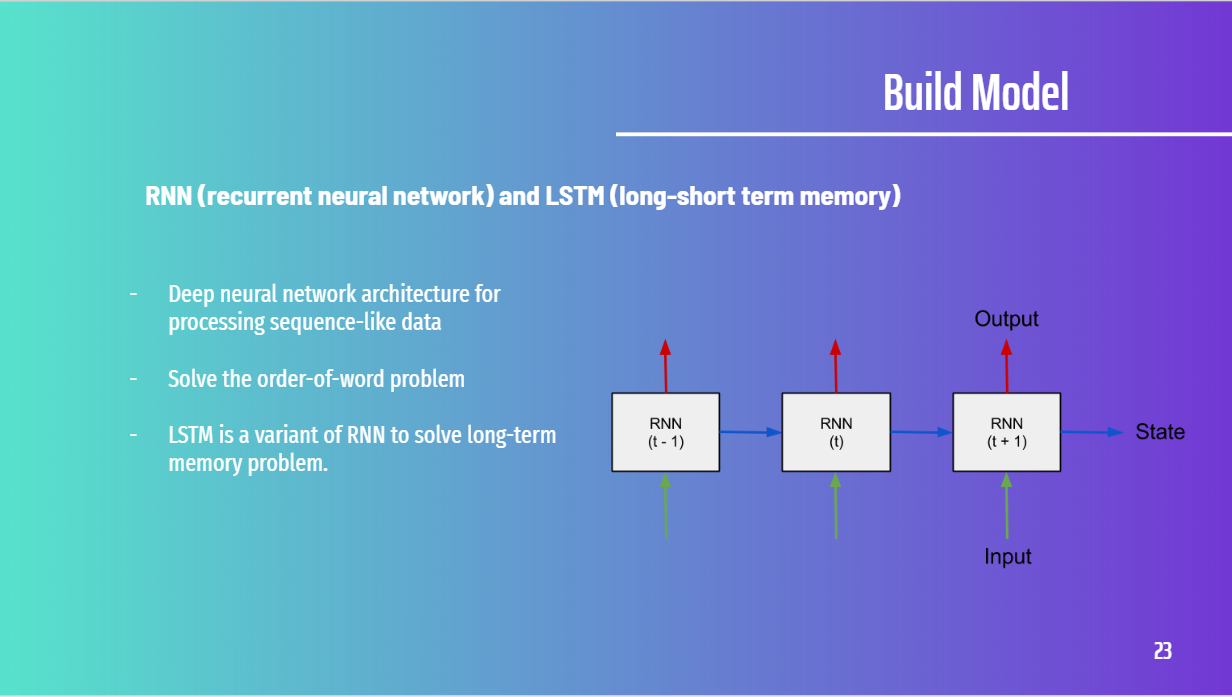
Word2vec helps find a custom-sized vector represent the words based on other words around them.

Thus, word2vec method well handle BoW's problem of storing large storage vectors and present the correlation between words better.



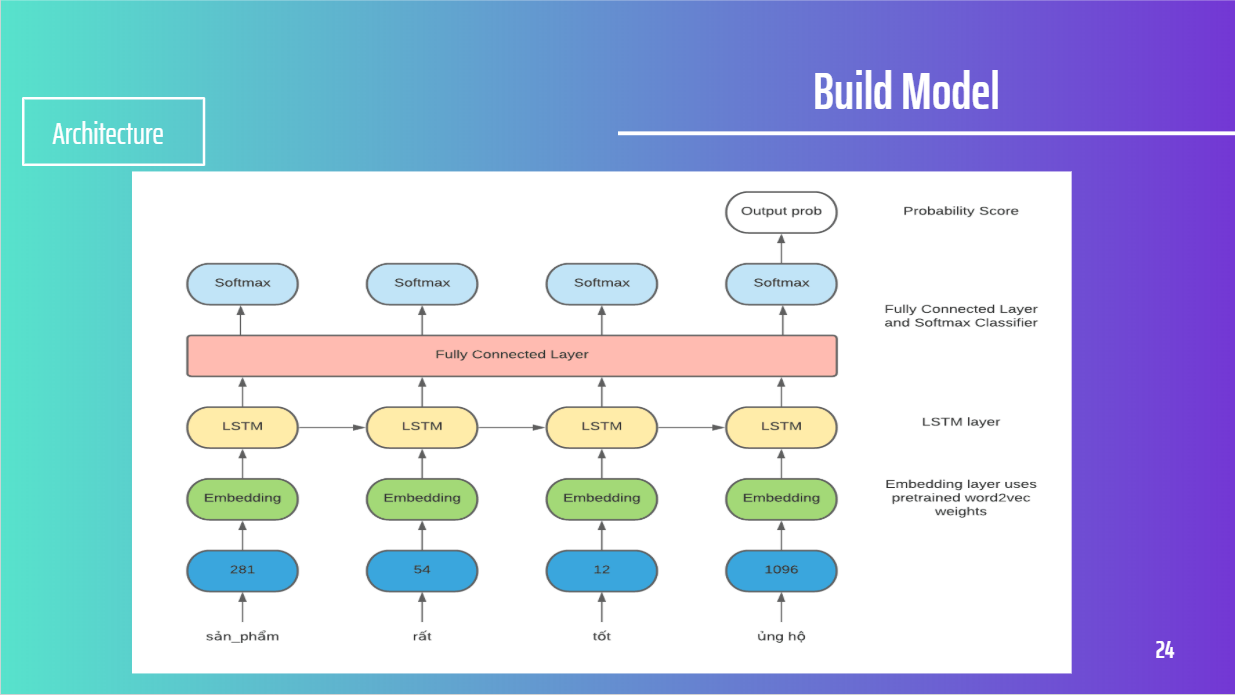
*Figure 22: Step 2: Processing - Result of Word2vec*

To acquire this result, we used the framework provided by gensim and trained it from scratch on a large corpus called VNTQ corpus which contains about 1.7 millions sentences collected from several Vietnamese electronic media site (baomoi, vnexpress,ect). We use the vectors got from word2vec model as the weights for embedding layer in our model.



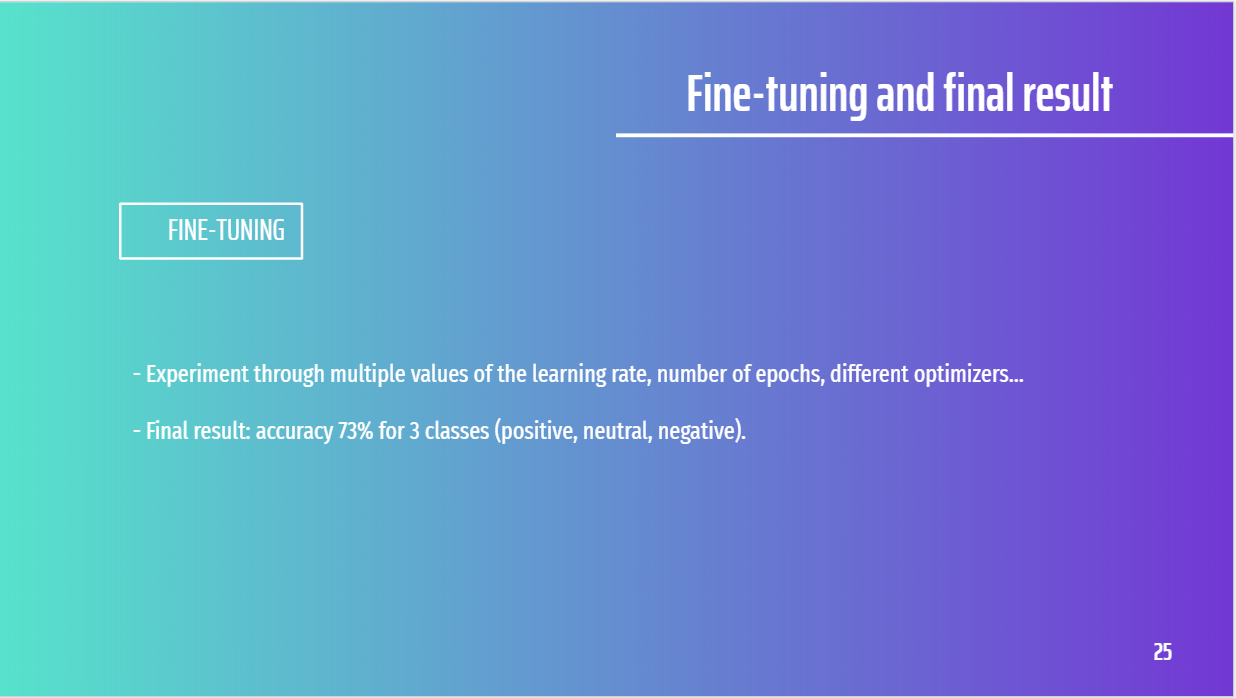
*Figure 23: Step 3: Build Model - RNN*

Step 3: Build model. As to solve the order-of-words problem, we used Recurrent Neuron Network (RNN). It's a deep architecture to deal with sequence-like data, that's just all we need to solve our remaining problem. In our case, we used Long-Short Term Memory - a variant of RNN to solve the vanishing gradient.



*Figure 24: Step 3: Build model - Architecture*

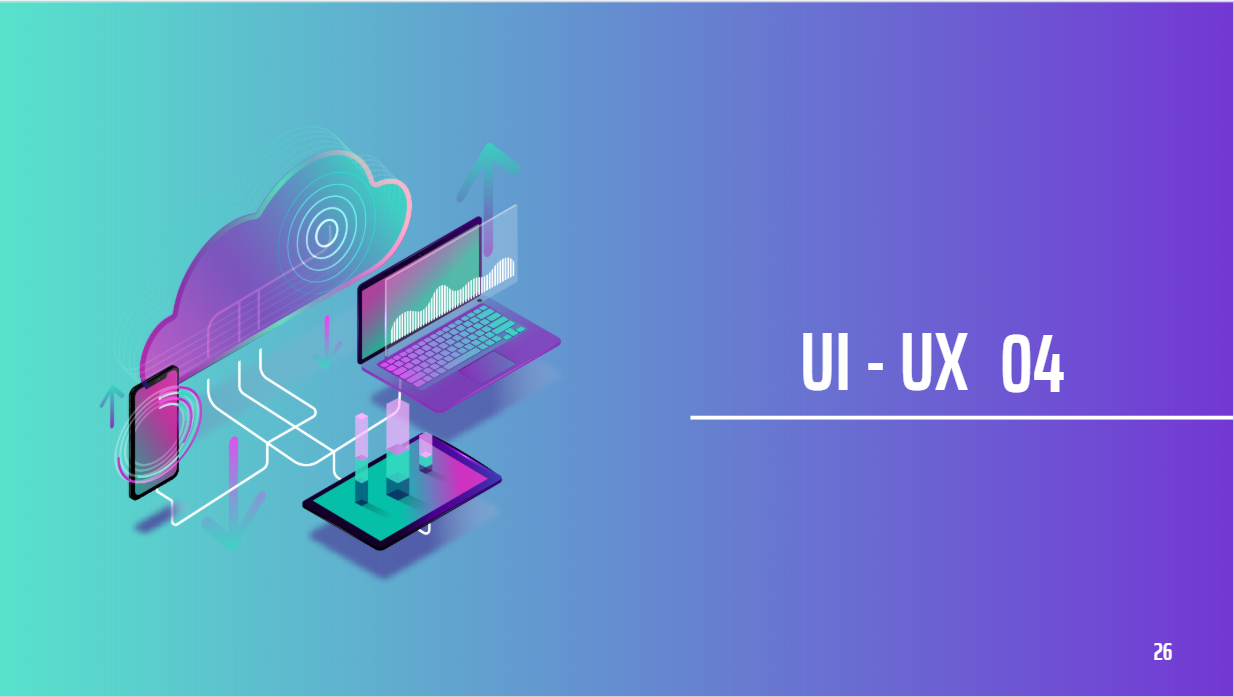
Our model include 3 layers: Embedding layer uses pre-trained word2vec weights, LSTM layer, Fully connected layer and softmax classifier.



*Figure 25: Step 4: Fine-tuning and final result*

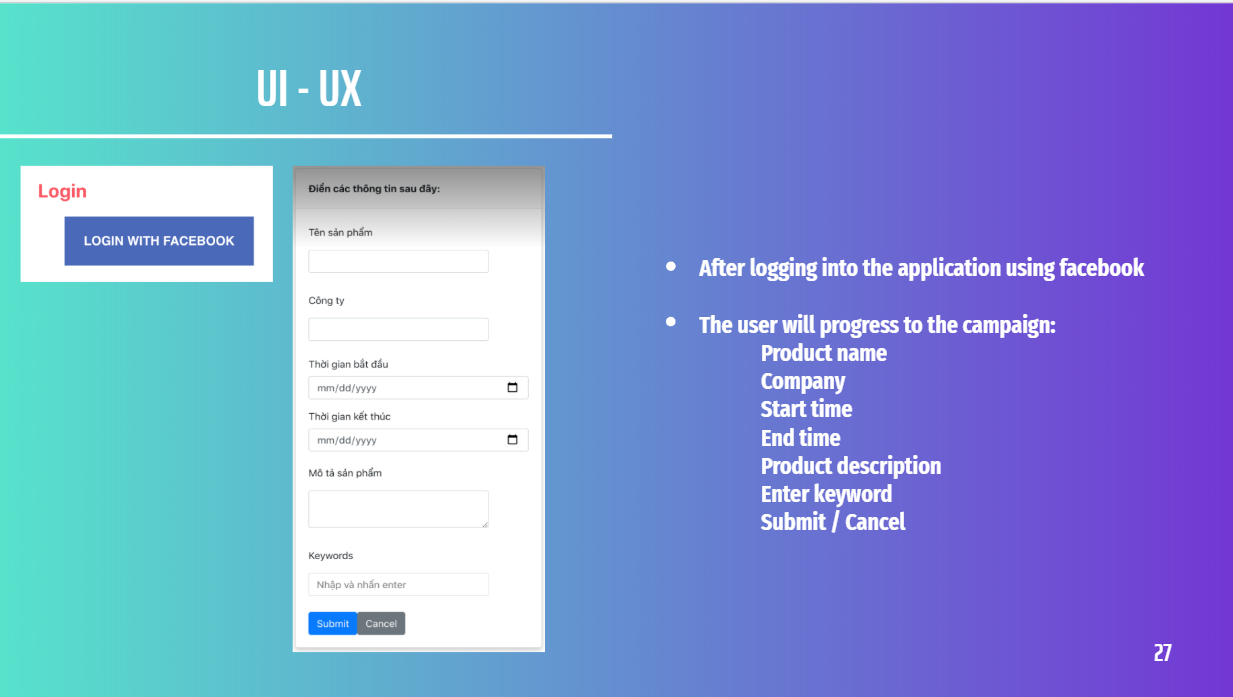
Step 4 – the final step is fine-tunning, and final result. In this step, we configured the model with different values of learning rates, number of epochs, and different optimizers, and got the final result: 73% accuracy for 3 classes (positive, neutral, negative).

## **4/ UI – UX**



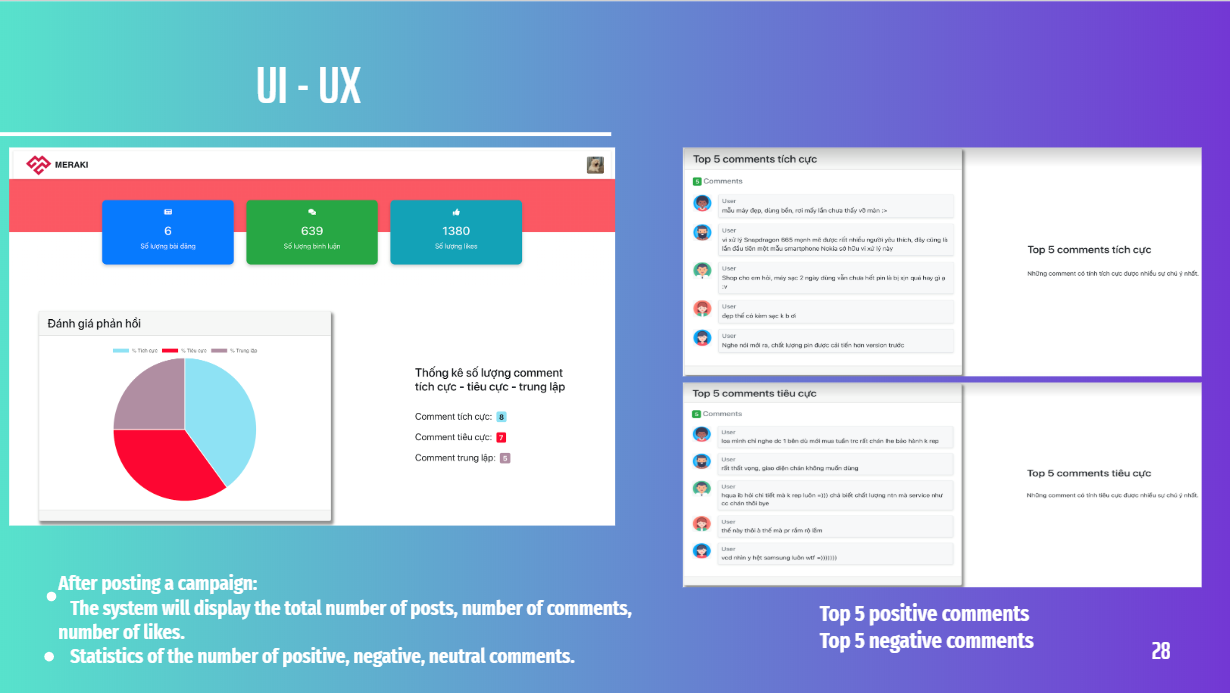
*Figure 26: UI - UX*

This is about our UI – UX.



*Figure 27: UI - UX - Login and Create a campaign*

After logging in with Facebook, users will create a campaign form.



*Figure 28: UI - UX - The total number of posts, comments, likes – Statistic of positive, negative and neutral comments - Top 5 negative and positive comment*

Then, user will get results displayed on the website about the number of posts that the user has posted, the number of comments, the number of likes. The chart shows the number of positive, negative, and neutral comments right after the results of the number of posts, comments, and likes you have obtained. Next is the top 5 positive comments and top 5 negative comments.



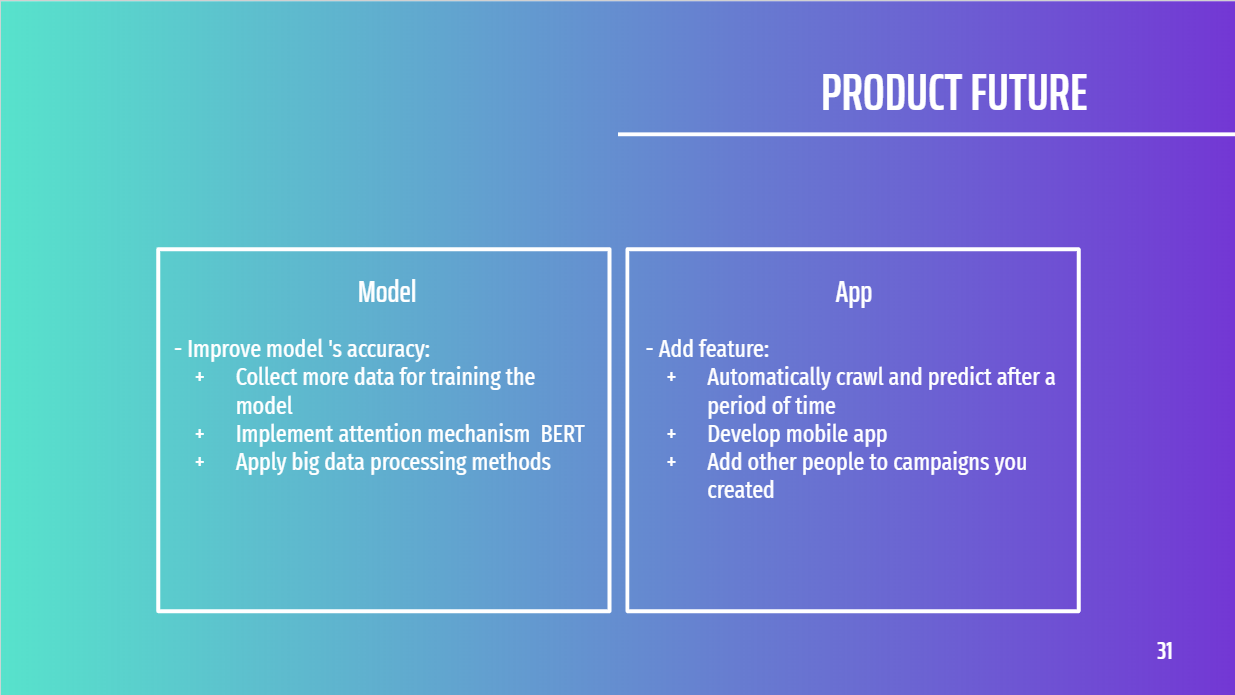
*Figure 29: UI - UX - Comment summary table and re-label*

Finally, in the summary table of comments, and especially in this table, users can relabel to each comment if detecting errors.

## **5/ Product future**



*Figure 30: Product future*



*Figure 31: Product future - Model and App*

In this section, we will develop our product in the future with model and app. Firstly, model accuracy improvement which will collect more data for training the model, implement attention mechanism and apply big data processing methods. Secondly, feature addition which we automatically crawl and predict after a period of time, develop mobile app and add other people campaigns you created.

## **6/ Struggle**



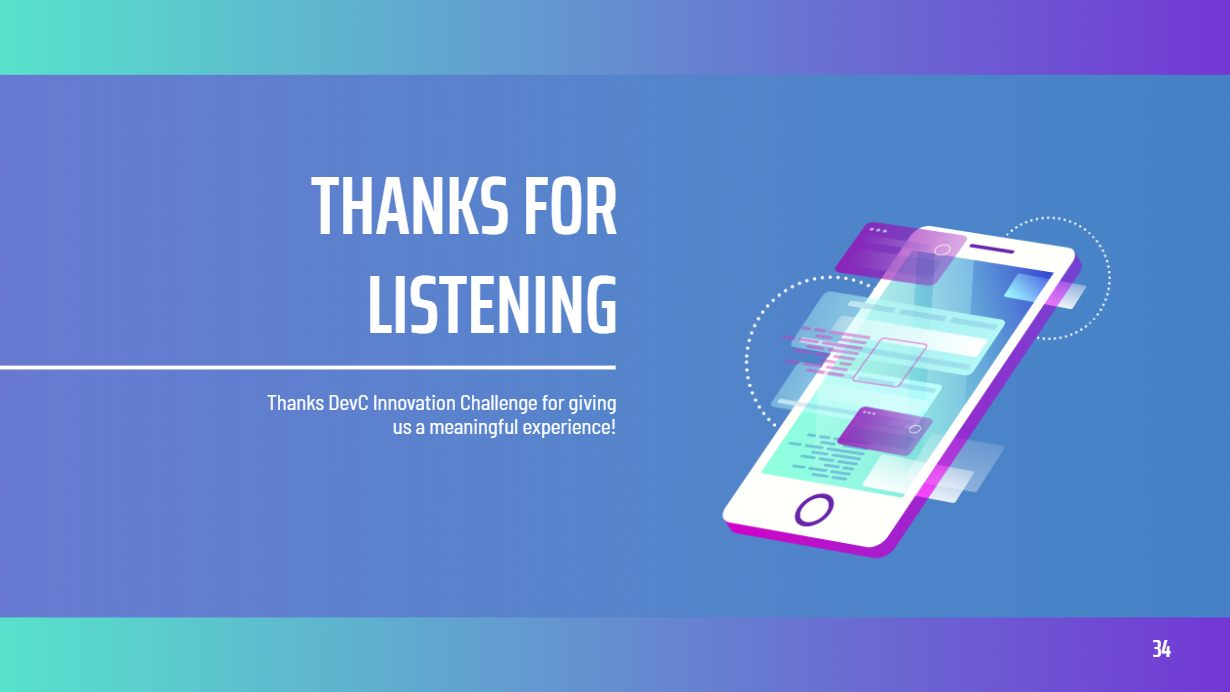
*Figure 32: Struggle*

In this final section: Struggle. These are the difficulties that teams encounter in terms of technical and communication.



*Figure 33: Struggle - Technical & Communication*

The first struggle is technical, approaching was difficult at first due to the new members. The process of finding documents and sources also encountered many obstacles. After completing the product, the interconnected parts of the product error and had to be fixed. The second struggle is the communication, we also meet some obstacles such as: We have not worked well together due to not understanding each other, we are not actively contributing and developing products due to the lack of connection to the team.



*Figure 34: Thanks for listening*

Thank you for your listen.

Thanks DevC Innovation Challenge for giving us a meaningful experience in student’s life.

# **VI/ REFERENCES**

<https://viblo.asia/p/so-luoc-word2vec-6J3ZgBoAKmB> [Accessed 24 November 2020].

<https://nttuan8.com/bai-13-recurrent-neural-network/> [Accessed 24 November 2020].