

**INTERNSHIP REPORT**

A Project Report Presented to

**Department of Information Technology**

Nguyen Tat Thanh University

NTT Institute of International Education

In Partial Fulfillment of the Requirements

For Internship Course

**March 18th, 2024 – May 19th, 2024**

**By**

1. **Nguyen Thai Minh Thien – 2000004762**
2. **Pham Trung Tin – 2000002491**
3. **Vo Minh Toan - 2000004394**

[**thienng.it@gmail.com**](mailto:thienng.it@gmail.com)

[**phamtrungtin5824@gmail.com**](mailto:phamtrungtin5824@gmail.com)

**Employer Information**

**Name of Company: Hinnova Company**

**Name of Supervisor: Mr. Huynh Nguyen Khac Huy**

**Phone: 0938915136 - Email: huynhnk@hinnova.com.vn**

**ABSTRACT**

This abstract explores our journey as an Software Developer Intern at Hinnova Company. Under the guidance and direction of Mr. Huy, we were not only equipped with useful knowledge but also gained extensive experience in practical work from a business perspective.

We were able to learn and experience knowledge and experience in Artificial Intelligence, Machine Learning, Deep Learning, deep learning training models, Python libraries, and the latest frameworks in natural language processing such as Bag-of-Words (BoW), TF-IDF, Word2Vec, GloVe, FastText, Attention Mechanism, Transformer, BERT, GPT, NER, …

On behalf of the interns, we would like to extend our heartfelt gratitude to the company for providing us with the opportunity to learn, work, and experience a professional and attentive work environment.

We are particularly grateful to Mr. Huy for his unwavering support and guidance throughout our internship, which has made it an invaluable experience.

We also express our sincere appreciation to our instructors for creating the opportunity for us to intern in such a supportive environment and for providing us with a stepping stone for our future endeavors by connecting us with experienced mentors and seniors.

From the bottom of our hearts, we extend our warmest wishes to the esteemed company, the esteemed school, our respected teachers, and our esteemed seniors for an abundance of good health, continued success in your endeavors, and unwavering peace in your lives.

**By**

Nguyen Thai Minh Thien – 2000004762

Vo Minh Toan - 2000004394

Pham Trung Tin - 2000002491

**TABLE OF CONTENTS**

[1.](#_heading=h.gjdgxs) Job Background 1

[1.1.](#_heading=h.30j0zll) Organization 1

[1.2.](#_heading=h.1fob9te) Project Management 1

[1.3.](#_heading=h.3znysh7) Development Environment 1

[2.](#_heading=h.2et92p0) Accomplishments 2

[3.](#_heading=h.tyjcwt) Technical Detail 3

[3.1.](#_heading=h.3dy6vkm) Requirements 3

[3.2.](#_heading=h.1t3h5sf) Design Process 3

[3.3.](#_heading=h.4d34og8) Design Detail 3

[3.4.](#_heading=h.2s8eyo1) Implementation Detail 4

[3.5.](#_heading=h.17dp8vu) Testing 4

[4.](#_heading=h.3rdcrjn) Lessons Learned 5

[5.](#_heading=h.26in1rg) References 6

**LIST OF TABLES**

**No table entries found.**

**LIST OF FIGURES**

**No table of figures entries found.**

**ACKNOWLEDGEMENTS**

*When I worked here, I seemed to be happy because of friendly and professional people. They taught me how to do research on everything, how to write a document, how to do a project in IT, and how to work with teams. It’s such invaluable moments that I’ve learned.*

- Nguyen Thai Minh Thien

# Job Background

Established in 2016, HINNOVA is a software development company with a team of highly qualified, professional, and experienced software professionals.

HINNOVA's Core Business Areas:

* Software development across various platforms: .Net, Java, Mobile, and more.
* Providing high-quality software personnel to strategic partners: Sacombank, Viettel, VIB, VietInfo, and others.
* Consulting on software architecture and solution design.

HINNOVA fosters a dynamic, creative, and comprehensive environment that promotes the overall growth and excellence of each individual member. We consistently reward our members with both tangible and intangible benefits commensurate with their contributions.

# Organization

Picture 1.1‑1: Hinnova’s Organization Chart



Picture 1.1‑2: [Hinnova’s workplace](https://glints.com/vn/companies/cong-ty-co-phan-hinnova/d8962c23-8096-4b1d-9e26-0fff26a9bafc)



Picture 1.1‑3: [Hinnova’s event](https://itviec.com/nha-tuyen-dung/hinnova)

Here is Hinnova website URL: [hinnova.com.vn](http://hinnova.com.vn/)

Here is Hinnova address: Lầu 7, 181 Cao Thắng, Phường 12, Quận 10, Tp.HCM, District 10, Ho Chi Minh

# Project Management

Explain how system requirements were gathered communicated at the outset, e.g., design review, customer meetings, group or staff meeting, etc, even if you were not part of that process. Continue on explaining what observable management habits or patterns were used to keep the project on track. Describe, to the best of your ability, the software project lifecycle management mechanisms.

Picture 1.2‑1: Software Development Life Cycle

# Development Environment

There are some Development Environment: Python IDE, Google Colab, Anaconda Navigator, Visual Studio Code, Pycharm Community Edition 2023, etc.

# Accomplishments

Provide an assessment of whether your internship allowed you to achieve your stated educational and career objectives and aspirations. Explain your assessment. That is, why and how you arrived at the stated assessment. For example, “My internship assignment allowed me to achieve my educational objective to learn more about and gain experience with the practice of software quality assurance. This was accomplished as a result of successfully completing my assigned task to apply commercial and in-house testing software to the company’s network management software modules.”

|  |  |
| --- | --- |
| **Week** | **Content** |
| 1 | 1. Research about AI (Artificial Intelligence). Definition, Real-world applications, many popular fields in AI. 2. Machine Learning Fundamentals. Understand the basics definition of Machine Learning such as Supervised Learning, Unsupervised Learning, Reinforcement Learning, and other popular algorithm: Linear Regression, Logistic Regression, Decision Trees, and Naïve Bayes. |
| 2 | 1. Reasearch about Deep Learning: neural networks, DL Structure such as CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), and Transformer. |
| 3 | 1. Research Python in basics 2. Google Colab tool 3. CPU, GPU, TPU, etc. 4. Research NLP fundamentals:  * Train and Test Data * Bag-of-Words and TF-IDF * Word Embeddings: Word2Vec, GloVe, and FastText, etc. * Categorization |
| 4 | 1. Research about Attention Mechanism 2. Transfer Learning: Pre-trained models such as BERT, GPT (Generative Pre-trained Transformer), etc. |
| 5 | 1. Research about NER (Named Entity Recognition):  * Step-by-step NER process. * State-of-the-art NER models * State-of-the-art Vietnamese NER models. |
| 6 | 1. Coding:  * Find Source Code about popular NER models 🡪 Research 🡪 Setup 🡪 Launch. * Indicate a Real-world NER application. |
| 7 | 1. Coding:  * Find Source Code about popular NER models 🡪 Research 🡪 Setup 🡪 Launch.   Indicate a Real-world NER application. |
| 8 | 1. Coding:  * Find Source Code about popular NER models 🡪 Research 🡪 Setup 🡪 Launch.   Indicate a Real-world NER application. |
| 9 | 1. Coding:  * Find Source Code about popular NER models 🡪 Research 🡪 Setup 🡪 Launch.   Indicate a Real-world NER application. |
| 10 | Final Report |

# Technical Detail

**The Challenge of Customer Feedback Overload**

In today's service-oriented world, organizations from online retailers to government agencies bombard us with customer satisfaction surveys. This constant feedback loop can generate a massive amount of data – potentially overwhelming for managers to analyze manually.

**Enter Machine Learning and Text Analytics**

Fortunately, advancements in Machine Learning offer a solution. A specific area called Natural Language Processing (NLP) allows machines to understand human language. Combined with Text Analytics, NLP powers applications like Feedback Summarizers.

**Feedback Summarization: Saving Time and Extracting Insights**

Feedback Summarizers automatically process and condense customer feedback, extracting key points while preserving the original meaning. This not only saves time but also provides managers with valuable insights previously buried within vast amounts of text data.

**NLTK: Your Gateway to Powerful NLP with Python**

If you're interested in diving deeper into Data Analytics, NLP holds immense potential. Python, a popular programming language, offers robust NLP libraries like NLTK (Natural Language Toolkit). We'll utilize NLTK throughout this exploration to unlock the power of text analysis and NLP for customer feedback.

# Requirements

Before we deep-dive into the project, firstly, install NLTK module on our system with following commands:

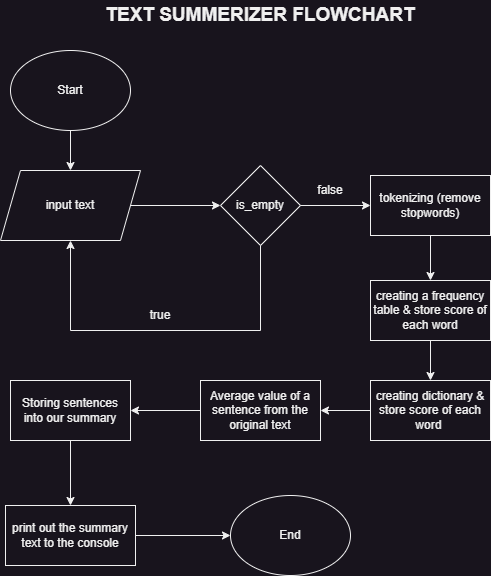
**pip install nltk**

Also, we need to meet some requirements:

* NLTK (3.8.1)
* Python (3.11.9) (64-bit)
* Microsoft Visual Studio (1.89.1)

# Design Detail

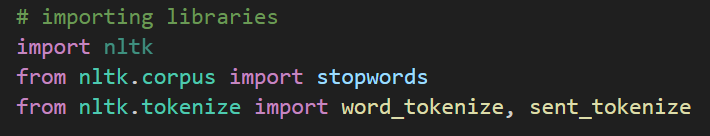
Firstly, here is our flowchart:



Picture 3.2‑1: Text-summerizer flowchart

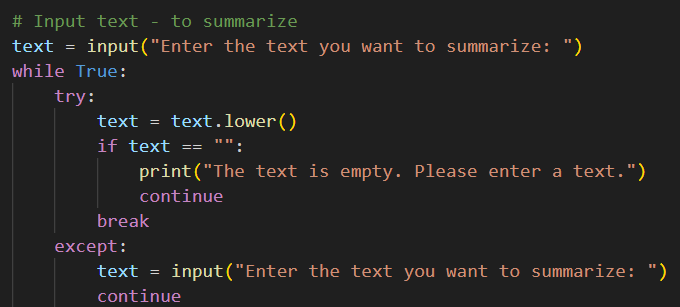
Secondly, we let’s kick off the project:

**Step 1: Import the libraries**



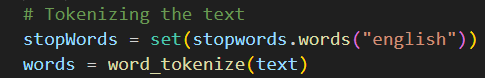
Picture 3.2‑2: Importing libraries

**Step 2: Input text – to summerize**



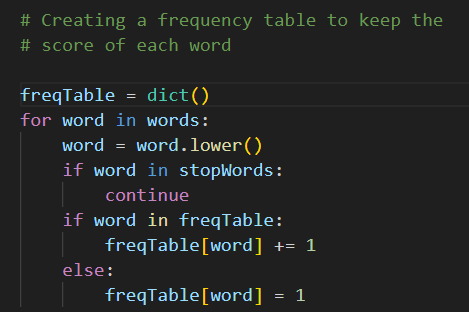
Picture 3.2‑3: Try to input text with some conditions

**Step 3: Tokenizing the text**



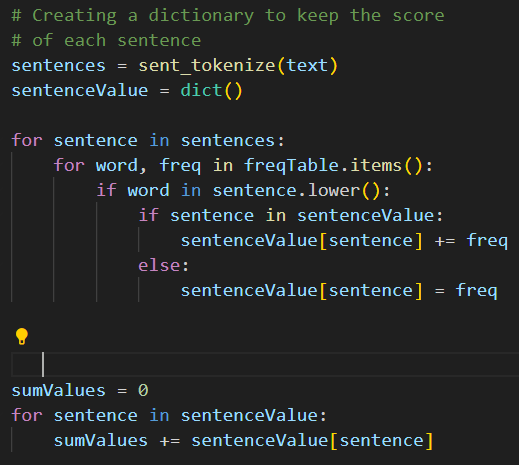
Picture 3.2‑4: Tokenizing the text

**Step 4: Creating a frequency table to keep the score of each word**



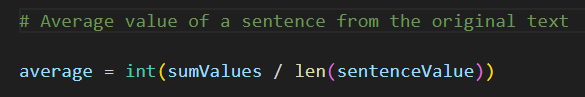
Picture 3.2‑5: Frequency table

**Step 5: Creating a dictionary to keep the score of each sentence**



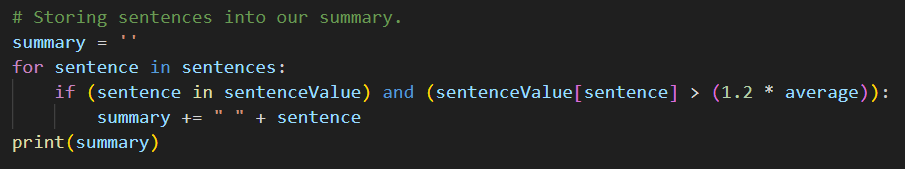
Picture 3.2‑6: Dictionary

**Step 6: Average value of a sentence from the original text**



Picture 3.2‑7: Average value

**Step 7: Storing sentences into our summary**



Picture 3.2‑8: Store sentences

**Step 8: Print out**



Picture 3.2‑9: Print outto the console

# Testing

**Here is input:**

There are many techniques available to generate extractive summarization to keep it simple, I will be using an unsupervised learning approach to find the sentences similarity and rank them. Summarization can be defined as a task of producing a concise and fluent summary while preserving key information and overall meaning. One benefit of this will be, you don’t need to train and build a model prior start using it for your project. It’s good to understand Cosine similarity to make the best use of the code you are going to see. Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. Its measures cosine of the angle between vectors. The angle will be 0 if sentences are similar.

**And output:**

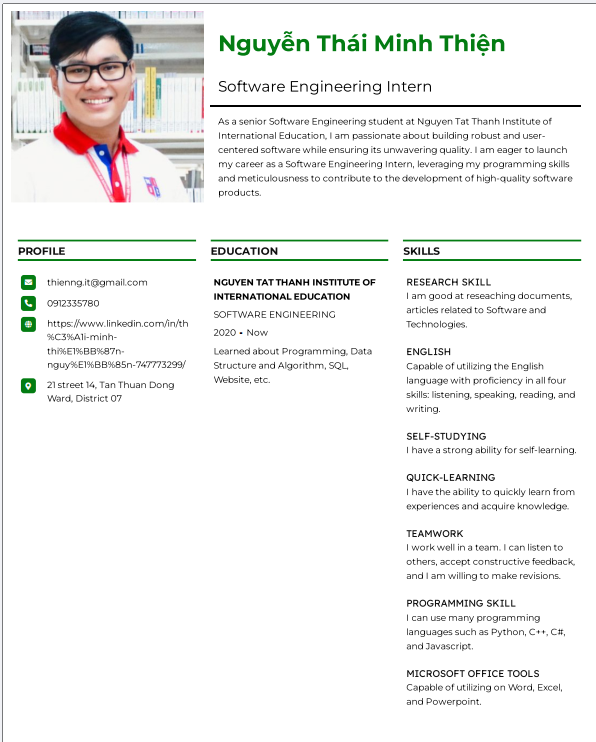
there are many techniques available to generate extractive summarization to keep it simple, i will be using an unsupervised learning approach to find the sentences similarity and rank them. cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

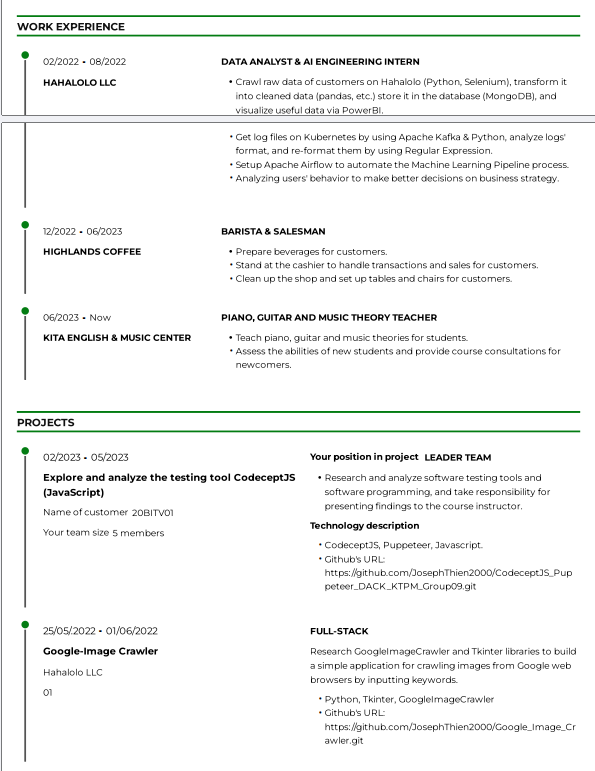
# Lessons Learned

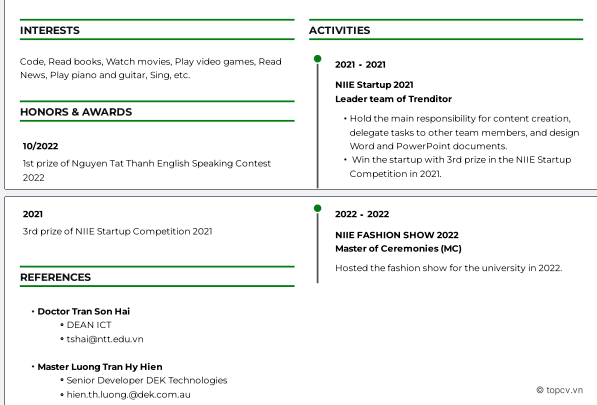
I learned how to research, setup and launch projects regarding NLP python libraries with multiple NLP frameworks and models such as BERT, NER, Bag-of-Words, TF-IDF, Word Embeddings, Attention Mechanism, Transformer, etc. and fundamental definitions about AI, Machine Learning, Deep Learning such as RNN, CNN, Linear Regression, Naïve Bayes, Transformer, etc.

Furthermore, I gained some soft skills when I worked here such as teamworking, Self-learning, Responsibility with tasks, how to wear polite clothes in workplace, communication skills when I worked with co-workers and so on.

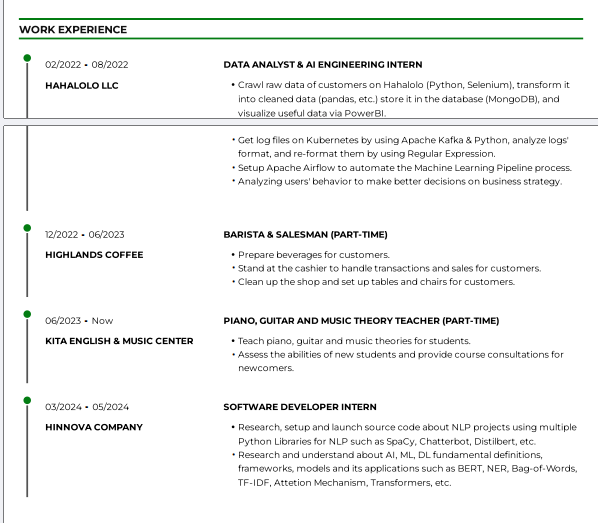
Job Skill requirements: Python, Natural Language Processing (NLP), AI (Artificial Interlligence), Machine Learning (ML), Deep Learning (DL), Bidirectional Encoder Representtations from Transformers (BERT), Named-Entity Recognition (NER), Attention Mechanism, Google Colab, CPU, TPU, GPU, github, etc.



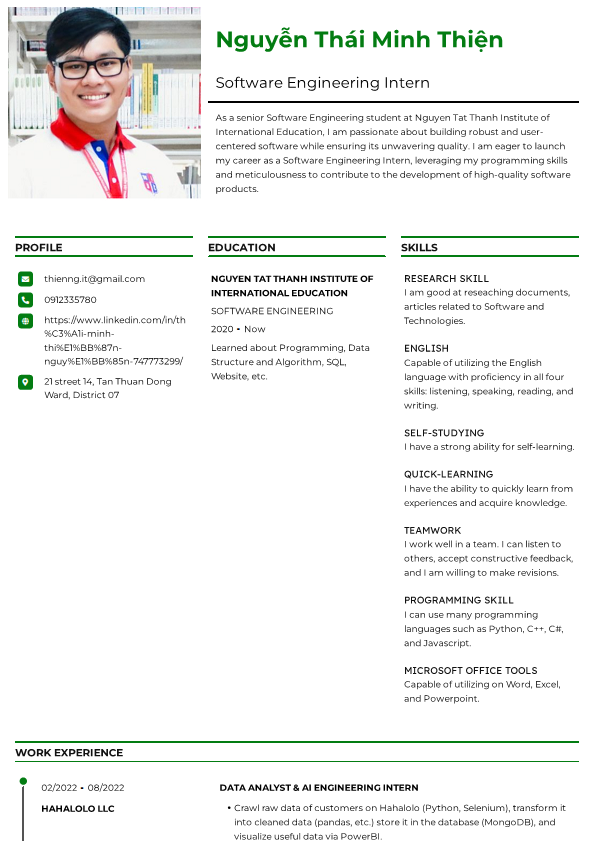




Picture 4‑1: Old CV



Picture 4‑2: New CV (edited part)



# References

Admin. (2023, 03 28). *DistilBERT: Smaller, Faster, and Lighter BERT Model (with Python Examples)*. Retrieved from www.pythonprog.com: https://www.pythonprog.com/distilbert/