

# COMP4801 Final Year Project

## Project Plan

Prepared By:

Richard Bryan

Md Abdullah Al Mahin

Sabyasachi Purkayastha

Yuen Yau Yoo

# 1. Table Of Contents

<b>1. Table Of Contents</b>	<b>2</b>
<b>2. Introduction</b>	<b>3</b>
2.1 Background & Problem Statement	3
2.2 Objectives	3
2.3 Outline	3
<b>3. Related Products &amp; Literature Review</b>	<b>4</b>
<b>4. Product &amp; Features</b>	<b>4</b>
4.1 Basic Features	4
4.2 Artificial Intelligence & Machine Learning (AI/ML) Features	5
4.3 Final Product	5
<b>5. Project Deliverables</b>	<b>6</b>
5.1 Multi-Page Frontend	6
5.2 Extensible Back-end API for the Website	6
5.3 Centralized NoSQL Database	6
5.4 Database Querying Interface	7
5.5 Multi-Source Data Collection MicroService	7
5.6 Dedicated Machine Learning Operations System & API	7
5.7 User Behaviour & System State Logging System	7
<b>6. Technological Framework</b>	<b>8</b>
6.1 General Programming Languages	8
6.2 Front-end & User Onboarding	8
6.3 Back-end APIs	9
6.4 Database Management & Type Safety	9
6.5 Version Control	10
6.6 Machine Learning	10
<b>7. System Infrastructure</b>	<b>11</b>
<b>8. Challenges</b>	<b>11</b>
8.1 Frontend & UI/UX	11
8.3 Backend Services	12
8.3 Machine Learning	12
<b>10. Project Management</b>	<b>13</b>
<b>11. Provisional Project Schedule</b>	<b>14</b>
<b>12. Conclusion</b>	<b>15</b>

## **2. Introduction**

### **2.1 Background & Problem Statement**

Course planning and registration are integral components of a university student's academic journey, bearing profound significance in shaping their educational experience and future career prospects. The process involves selecting courses that align with their academic interests and career goals whilst fulfilling requisite criteria, all within the constraints of course availability and scheduling preferences. Efficient and informed decision-making in this regard can significantly impact a student's academic success and overall satisfaction with their university experience.

However, this process often proves to be intricate and overwhelming, especially when students must navigate a complex web of course offerings, prerequisites, schedules, and personal preferences. For the University of Hong Kong (HKU) in particular, there is an absence of a centralized course database offering important course information such as course and professor reviews. Consequently, students generally rely on word-of-mouth information from their network of HKU students or alumni. Furthermore, due to the lack of centralization of course information, students often find themselves scouring the internet to consolidate information regarding courses offered by various faculties or departments. In light of this, there is a pressing need for innovative solutions that harness the power of technology and machine learning to streamline course planning and registration, empowering students with personalized, data-driven recommendations and scheduling options. This underscores the importance of our Final Year Project (hereinafter referred to as FYP), which aims to address this fundamental challenge by creating a user-centric course management platform equipped with intelligent features.

### **2.2 Objectives**

The primary objective of our FYP is to develop a user-centric course management platform that leverages machine learning and data-driven insights to simplify the course planning and registration process for HKU students. This platform will act as a centralized course database, with the aim of offering a one-stop solution to all course-related questions that HKU students may come across. It will provide essential course information such as course descriptions and schedules, as well as in-depth reviews addressing critical course aspects such as usefulness, workload, and teaching by professors to assist students in making informed decisions. Apart from that, we plan to implement scheduling functionality that integrates seamlessly with the course platform to alleviate students from the tedious manual resolution of timetable conflicts between their desired courses for a particular academic semester. Furthermore, we seek to incorporate Artificial Intelligence & Machine Learning (AI/ML) capabilities into our platform to empower students with personalized course recommendations based on several factors such as their academic history, interests, and career goals. Ultimately, we aim to enhance the overall experience for university students by making the course selection and planning process more intuitive, efficient, and personalized.

### **2.3 Outline**

The remainder of this paper proceeds as follows. First, we will perform analyses on related platforms that are currently in the market to obtain an understanding of competing products. We then present proposed features of our platform that would allow us to differentiate ourselves from our competitors. Next, we go through some deliverables for a more tangible and measurable overview of the phases of our project. With our product and deliverables in mind, we then move on to describe the proposed technological framework as well as the system architecture that we will employ and adhere to throughout the entire development process. Following that, we explore several challenges that we

foresee ourselves facing. We then briefly describe the project management aspect of our FYP, which addresses the distribution of roles and responsibilities amongst our group members, and also provide a tentative schedule regarding the expected completion dates of the discussed project deliverables. Finally, we close with a recap of the problem statement and primary objective of our project, and an address of several key concerns.

### 3. Related Products & Literature Review

From our research, we have only found one other competing product in the market, namely *RIC* 杂货铺, accessible through the website <https://richku.com/>. This platform was developed in recent years by the Rights and Interests Committee of the Mainland Undergraduate Association of HKU (香港大学内地本科生联合会权益保障组). Given the fact that all information on the website is only available in Chinese, we conclude that it caters specifically to a Chinese-literate user base.

We will analyze the course selection platform, 选课平台, offered on their website. As of today, it serves three main purposes. Firstly, it allows students to query basic information regarding a course such as its description, offering department, timetable, and past grading statistics. Secondly, it provides access to course reviews written by students. These course reviews contain useful information such as the academic year and semester that the student took the course in, the teaching professor, and its helpfulness measured by a like and dislike metric. Thirdly, it offers a scheduling feature, enabling students to schedule courses for the upcoming academic year and optionally, to export their schedule as a *.ics* file.

In comparison to our proposed product, we want to highlight two main distinctions: target user base and AI/ML capabilities. In contrast to *RIC* 杂货铺's target user base who seem to be Chinese-literate, we aim to develop an inclusive platform catered towards all HKU students regardless of their nationality or linguistic capabilities. Furthermore, we plan to integrate AI/ML features such as a personalized course recommendation engine, which are currently not found in *RIC* 杂货铺's course selection platform.

## 4. Product & Features

### 4.1 Basic Features

Our Final Year Project (FYP) aims to deliver a user-centric web application for course management, powered by AI/ML capabilities. The core functionality of our platform revolves around a centralized course database, providing HKU students with easy access to essential course information. By consolidating course descriptions, prerequisites, co-requisites, anti-requisites, and timetables, we eliminate the need for students to navigate multiple faculty or department websites. Our goal is to offer a comprehensive, one-stop solution for all inquiries related to courses offered by HKU.

In addition to the course database, our platform will feature a reliable review system for both courses and professors. Students will have the opportunity to review courses based on various criteria, including usefulness, workload, and difficulty. Similarly, professors will be evaluated on factors such as clarity and engagement. To ensure the integrity of these reviews, we will implement a verification system that rigorously analyzes and validates all published reviews. By providing verified course and professor feedback from past students, we empower current students to make informed decisions during the course planning process.

Another key aspect of our project is the development of an intuitive scheduling feature that seamlessly integrates into the course planning workflow. After selecting their desired courses, users

can explore our scheduling page. Our aim is to create an automatic scheduling function that optimizes the selection of subclasses for each course, taking into account user preferences such as morning or evening classes, and packed or relaxed class days. Additionally, we plan to incorporate an export feature that allows users to export their schedules as ICS files. This universal calendar format is compatible with popular third-party calendar apps like Google and Outlook Calendar, providing users with greater flexibility and convenience.

## **4.2 Artificial Intelligence & Machine Learning (AI/ML) Features**

Machine learning plays a pivotal role in enhancing the user experience on our platform by leveraging its capabilities in various aspects. One significant application is spam detection, which effectively filters out irrelevant reviews, ensuring that students only encounter genuine and high-quality feedback. By implementing this feature, we build trust in the platform's recommendations, empowering students to make well-informed decisions regarding their course selections.

Moreover, employing natural language processing (NLP) techniques allows us to extract key insights from reviews, providing students with concise yet valuable assessments of courses and professors. By summarizing main points, highlighting pros and cons, and presenting essential details, we help students quickly evaluate whether a course aligns with their needs and goals. To personalize the student experience, recommendation engines can take into account factors such as academic history, interests, career goals, and the popularity of courses and professors. This enables us to suggest personalized course recommendations tailored to each student's specific requirements. Through optimizing course selection, we aim to enhance academic performance and overall satisfaction.

Sentiment analysis on reviews further facilitates efficient comparisons by summarizing the overall sentiment towards courses and professors. Students can swiftly gauge the quality of options through sentiment scores, thereby aiding their decision-making process. Algorithms can also analyze factors such as graduation requirements, prerequisites, workload, and availability to generate optimized recommendations for scheduling and planning. This personalized approach improves on-time graduation rates and reduces scheduling conflicts. Additionally, we can provide further filtering and recommendations based on students' past courses, prerequisites, and individual goals, refining the available options and reducing the complexity of course selection. By encouraging the exploration of related topics, we can suggest similar courses to broaden students' horizons and expand their academic horizons.

## **4.3 Final Product**

We aim to revolutionize the course management experience for HKU students by delivering an innovative platform. This platform optimizes the academic journey by consolidating a centralized database of essential course information and reviews, creating a reliable and trusted resource. Utilizing AI-enabled spam detection and natural language processing, we extract valuable insights from reviews to provide informed decision-making. Our platform offers personalized recommendations, intelligent scheduling, and adaptive filtering, all powered by machine learning. This ensures a tailored experience based on each student's academic goals and interests. By leveraging sentiment analysis, we enable efficient comparison of options, facilitating the decision-making process. With an intuitive interface, a comprehensive knowledge base of courses, and intelligent machine-learning capabilities, our platform empowers students to navigate course planning and selection seamlessly. Our user-centric approach, driven by AI and machine learning, empowers students to make well-informed decisions based on data, unlocking their full academic potential. Ultimately, we aim to redefine the course management experience by providing students with a platform that optimizes their academic journey.

## **5. Project Deliverables**

In any software engineering endeavor, the identification and definition of deliverables are paramount to project success. These deliverables are tangible manifestations of our progress and accomplishments throughout the course of our final year project. They provide a clear framework for assessing our achievements and ensuring that we are on track to meet our goals and objectives. In this section, we outline the comprehensive list of deliverables that will be produced, presenting a detailed overview of each milestone that will mark our journey toward the completion of this endeavor. Hence, we aim to complete the following deliverables.

### **5.1 Multi-Page Frontend**

The primary, user-facing deliverable of the project entails the creation of a multi-page frontend interface. Comprising seven distinct pages, we will aim to combine the latest trends in UI/UX design with the established look of the HKU ecosystem to deliver a platform that is both modern and conforming. To elaborate on the pages themselves, there will be a centralized course database page that simplifies navigation and access to course-related information with the aid of smart filtering and sorting options. Additionally, individual course information screens, featuring detailed content providing users with comprehensive insights, will also be established. The frontend will include a user profile page with the option of uploading personal information such as student transcripts, allowing users to tailor their experience to their preferences. To facilitate user interaction, there will also be a user login and registration experience page, ensuring a seamless onboarding process for new users. Finally, the design of our frontend will prioritize responsiveness, ensuring that users can access and interact with the platform smoothly across various devices, including desktop computers and mobile devices.

### **5.2 Extensible Back-end API for the Website**

In partnership with our front-end, the project also involves the development of an extensible back-end API intended to serve as the backbone of the primary website. The API will be designed to be lightweight, ensuring efficient performance. The key aim for this back-end API is a high level of customizability. With a look to the future, users should have the flexibility to adapt the API to their specific requirements through a range of configuration options. This customization capability will empower users to fine-tune the API's behavior and adapt it, enhancing versatility and applicability across different scenarios.

### **5.3 Centralized NoSQL Database**

Our systems will require the implementation of a robust, centralized NoSQL database system. The database system will be designed with scalability and flexibility in mind. To achieve efficient data storage, the database is optimized for handling various data types and structures, making it well-suited for diverse data sources. Its scalability ensures that it can seamlessly grow to handle increasing data volumes, avoiding bottlenecks or performance issues. Moreover, we will look to make sure that the related schema and relevant information stored conform to HKU's used database schemas, ensuring a smooth transition into the HKU ecosystem should the need to do so arise.

## **5.4 Database Querying Interface**

To further ensure a smooth transition between different databases, we would also look to develop an intuitive database querying interface. This interface will be designed to offer advanced search and filtering capabilities, making it easier for users to retrieve specific information from the database. One of the key features of this querying interface is its support for various query types. Users will have the flexibility to choose from different query options, empowering them to extract meaningful insights from the database. Whether it's simple searches, complex filters, or specific data retrieval, the interface will facilitate a wide range of query types to cater to diverse user needs.

## **5.5 Multi-Source Data Collection MicroService**

To saturate this database with relevant information, we would also need to create and deploy a versatile Multi-Source Data Collection MicroService. This service will be engineered to proficiently accumulate data from a wide array of sources and seamlessly integrate it into the project's central database. The system will also be able to supplement existing data should the database itself be shifted to another provider. To ensure precision and efficiency, the project will begin with a comprehensive definition of the data sources, encompassing APIs and user-generated content, while establishing a robust data ingestion process. The architecture of this data collection system is thoughtfully designed to operate in parallel with the primary API. This parallel architecture serves two vital purposes: first, it preserves the freshness of the data by continuously updating and synchronizing information from various sources, ensuring that users access the most up-to-date data available. Second, it contributes to data accuracy by cross-referencing and verifying information from multiple sources, enhancing the overall quality of the data stored in the project's database.

## **5.6 Dedicated Machine Learning Operations System & API**

To support artificially intelligent features of the website, we would set up a dedicated Machine Learning Operations API and develop individual machine learning models, each responsible for executing specific tasks outlined in the products and features section. To facilitate the seamless integration of machine learning into the project, a dedicated machine learning API server will be set up and meticulously configured. This API server will be responsible for powering various machine learning operations within the project. The tasks performed by the server encompass a range of functionalities, including recommendation algorithms, sentiment analysis, and predictive modeling. Each task will be meticulously specified, optimized, and integrated into the project to ensure not only their efficient operation but also their ability to provide valuable insights and functionality to users.

## **5.7 User Behaviour & System State Logging System**

Finally, there is a need to support both machine learning operations and enhance the user experience through the appropriate collection and storage of user behavior, while protecting user anonymity. Firstly, a comprehensive list of user actions and system events will be defined to capture relevant data. This involves identifying and specifying the various useful interactions users have with the platform, as well as significant system events that impact its operation. Secondly, to protect user privacy while still enabling effective data analysis, robust data protection and anonymization techniques will be employed. This ensures that personally identifiable information is safeguarded and that data is handled in compliance with privacy regulations. Anonymization techniques may involve removing or encrypting sensitive data elements.

## 6. Technological Framework

The selection of programming languages and frameworks is a pivotal decision that underpins the technical foundation of our software engineering final-year project. It forms the bedrock upon which our project's architecture, functionality, and scalability will be built. In this section, we delve into the strategic choices we have made with regard to programming languages and frameworks, providing insights into the rationale behind our selections and highlighting the synergies that these technologies bring to our development process. To facilitate the development of our deliverables, we will utilize the following technologies.

### 6.1 General Programming Languages

TypeScript is our preferred general programming language of choice for this project, and it serves as the foundation for many of the frameworks we will be using. TypeScript brings static typing to JavaScript, helping catch type-related errors during the development process, and providing an extra layer of safety and reliability to the codebase. It ensures that variables and functions are used consistently with their intended types. TypeScript also allows developers to define custom data structures using interfaces. Additionally, it provides type annotations that allow for the explicit definition of variable and object types. These annotations improve code clarity and serve as valuable documentation. Developers can easily understand the expected data shapes and how different components of the code interact with each other.

Python is a highly favored programming language for machine learning, primarily due to its wealth of libraries and frameworks tailored for this field, and consequently is our preferred language of choice for our machine learning operations. Python offers a robust foundation for machine learning. Its extensive list of libraries and frameworks, such as TensorFlow and PyTorch, provide essential tools for building, training, and deploying ML models. This ecosystem empowers developers to work efficiently and leverage state-of-the-art techniques in the machine learning domain. Moreover, Python introduces Pandas, a powerful library for data manipulation and analysis. This tool simplifies the preprocessing and exploration of datasets, streamlining the initial stages of the machine-learning pipeline. Pandas' capabilities enable data scientists and engineers to efficiently clean, transform, and gain insights from their data. In addition to its rich machine learning ecosystem, Python also supports Flask, a lightweight web framework ideal for deploying machine learning models as RESTful APIs. This capability enables real-time interaction between web applications and machine learning models, facilitating predictions and data processing on the fly. Flask's simplicity and versatility make it a valuable tool for integrating machine learning into web applications seamlessly.

### 6.2 Front-end & User Onboarding

We have opted to utilize Next.js as the cornerstone of our website construction. Several key features have driven our choice - first and foremost, Next.js provides robust support for Server-Side Rendering (SSR). This capability significantly enhances our website's performance and Search Engine Optimization (SEO) by generating web pages on the server and delivering them as fully rendered HTML to users, resulting in quicker page loads and improved SEO rankings. Next, automatic code splitting in Next.js streamlines and optimizes JavaScript and TypeScript code, ensuring that only the necessary code is loaded when users access specific pages. This not only accelerates initial page loading times but also enhances overall performance. Furthermore, Next.js offers the flexibility of both SSR and Static Site Generation (SSG). This versatility allows us to choose the most suitable method for different aspects of the project. SSG, for instance, generates



static HTML files during the build process, further enhancing page load speeds and minimizing server load for content that remains relatively unchanged.

In our Next.js application, we have chosen to implement authentication using NextAuth.js, an authentication library that offers built-in support for various authentication providers like OAuth, JWT, and more. The library is designed to seamlessly fit into Next.js projects, providing a straightforward setup process. Additionally, it offers customizable authentication flows for user login and registration, allowing us to tailor the authentication experience to our project's specific needs. Another significant advantage of NextAuth.js is its robust session management capabilities. It takes care of session management tasks, including secure token storage and automatic session handling. This simplifies the implementation of authentication in our Next.js projects, ensuring that user sessions are managed securely and efficiently.

### **6.3 Back-end APIs**

In the realm of API development, our choice is Fastify. It boasts a set of features that align perfectly with our project requirements. Foremost, Fastify is renowned for its high-performance capabilities. It stands out for its exceptional speed and efficient memory consumption. These attributes render it well-suited for handling a substantial volume of concurrent requests, ensuring that our API can deliver optimal responsiveness even under heavy loads. Moreover, Fastify places a strong emphasis on extensibility and developer-friendliness. Its architecture supports the incorporation of plugins and offers a straightforward API. This design philosophy makes Fastify highly adaptable and customizable to meet various project needs. Developers can easily extend its functionality, facilitating the integration of specific features and functionalities tailored to our project requirements.

For our machine learning operations API, we will adopt the use of Flask. Flask is a popular and lightweight web framework in the Python ecosystem, known for its simplicity and versatility. It serves as an excellent choice for developing web applications, particularly when considering its integration capabilities with various technologies. It offers seamless integration with various Python libraries and extensions, allowing developers to incorporate features like authentication, database connectivity, and session management effortlessly. This capability enhances the interoperability and extensibility of Flask-based web applications.

### **6.4 Database Management & Type Safety**

MongoDB is our chosen NoSQL database management system. It is renowned for its scalability, particularly its horizontal scalability. This characteristic allows us to efficiently distribute data across multiple servers, making it an ideal choice for managing vast amounts of data and handling high-traffic applications. As our data requirements grow, MongoDB can seamlessly expand to accommodate the increased load, ensuring that our databases remain performant and responsive. Moreover, MongoDB features a flexible schema, which stands in contrast to traditional relational databases with rigid schemas. This dynamic schema capability means that we have the flexibility to add or remove fields within our documents without causing disruption to the entire database structure. This adaptability is particularly valuable as it allows us to evolve our data structures and adapt to changing project requirements without significant overhead.

To enable type-safe, remote procedure call (RPC) systems within our TypeScript-based applications, we have chosen the tRPC framework. As a TypeScript-first framework, tRPC is explicitly tailored for constructing type-safe, remote procedure call systems. It excels at enforcing strong typing throughout both the client and server components, significantly reducing runtime errors and enhancing code reliability. One of its standout features is automated code generation. tRPC is capable of automatically generating client and server code based on a shared API definition. This automation

not only expedites the development process but also ensures consistency across the API, guaranteeing that client and server components remain in sync. Furthermore, tRPC is highly flexible and extensible. It empowers developers to define intricate APIs with ease, offering room for customization of client and server behavior. This flexibility makes it suitable for a wide array of applications and diverse APIs, accommodating a broad spectrum of project needs.

## 6.5 Version Control

GitLab is our preferred version control platform for several compelling reasons, all of which contribute to the efficiency and flexibility of our development processes. Firstly, GitLab stands out as an Integrated DevOps Platform. It consolidates a comprehensive suite of DevOps features into a single platform, encompassing Continuous Integration/Continuous Deployment (CI/CD), issue tracking, code review, and more. This integration streamlines our workflow by centralizing essential development and operations functions in one place, enhancing collaboration and productivity. Additionally, GitLab offers the valuable benefit of flexibility in deployment options. We can choose between self-hosted and cloud-based solutions, aligning their deployment method with their specific needs and preferences. This versatility ensures that GitLab is accessible and adaptable for organizations of varying sizes and technological infrastructures.

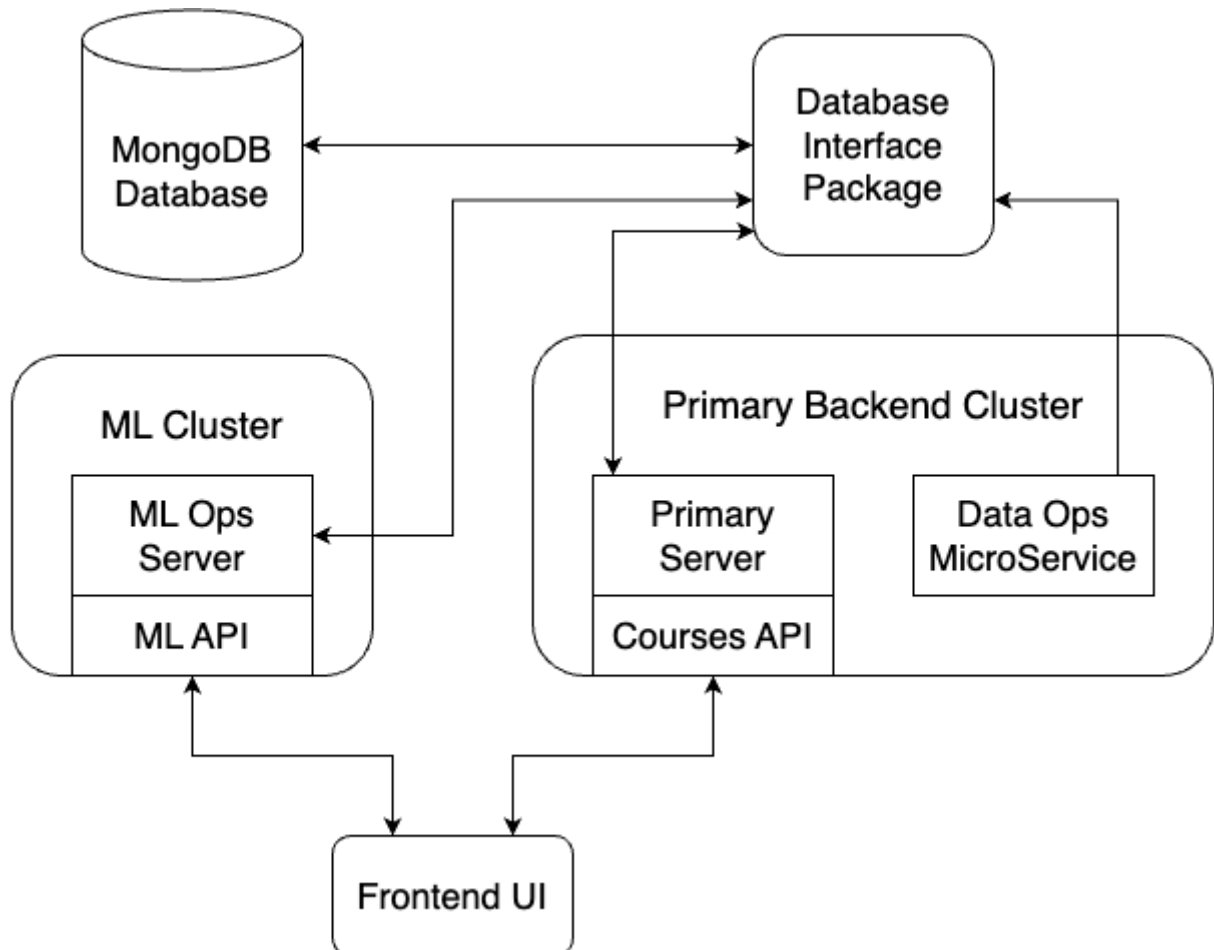
## 6.6 Machine Learning

In our machine learning (ML) and natural language processing (NLP) endeavors, we rely on a selection of specialized libraries and frameworks, each offering unique capabilities to support our projects. For machine learning, we turn to Scikit-Learn, a versatile and user-friendly library that excels in implementing a wide array of machine learning algorithms. Its intuitive API and extensive toolset make it a valuable resource for various ML tasks. Additionally, when it comes to deep learning models, including complex neural networks, we leverage TensorFlow and Keras, combining these two essential libraries to develop and train deep learning architectures effectively. Furthermore, PyTorch, known for its flexibility and dynamic computation graphs, serves as a powerful deep-learning library, empowering us to tackle complex tasks with ease.

In the realm of natural language processing, our toolkit includes NLTK (Natural Language Toolkit), a comprehensive library purpose-built for NLP tasks. NLTK offers a broad spectrum of functionalities, including text preprocessing, sentiment analysis, and text classification, making it an invaluable resource for NLP projects. Additionally, we will rely on spaCy, a high-performance NLP library known for its efficiency. It provides pre-trained models for various languages and excels in tasks like named entity recognition and part-of-speech tagging with exceptional accuracy and speed. This efficiency makes it our preferred choice for real-time NLP applications. Collectively, these machine learning and NLP libraries and frameworks enable us to address a wide range of challenges in these domains, ensuring the versatility and high capability of our projects.

## 7. System Infrastructure

The infrastructure deployed for our project constitutes the structural backbone that facilitates its seamless operation, scalability, and performance. The meticulous planning and implementation of this infrastructure are essential for ensuring the project's robustness and efficiency. In this section, we look at the intricacies of our infrastructure, elucidating the components, tools, and services that have been meticulously chosen to create a resilient and adaptable technological ecosystem for our project. With all features considered, the overall infrastructure is as follows.



## 8. Challenges

The success of any project relies heavily on effectively addressing various challenges throughout the data analysis process. These challenges encompass critical aspects such as data quality, model overfitting, etc. In this section, we will explore each challenge in detail, discussing their significance and the strategies employed to overcome them. By addressing these challenges head-on, our project will be well-equipped to deliver robust and trustworthy solutions.

### 8.1 Frontend & UI/UX

One significant hurdle could be the creation of a robust and intuitive front-end framework for timetable visualization. The presentation of a user-friendly interface that can seamlessly integrate with the complex back-end algorithms, while providing a visually pleasing and interactive timetable,

requires meticulous planning and execution. Addressing this challenge demands the creation of a dynamic front-end framework capable of handling real-time data manipulations and presenting them in an easily navigable manner, enhancing the user experience in course scheduling.

### **8.3 Backend Services**

A paramount concern in the development of this platform is the protection of student data privacy. Storing sensitive information necessitates the implementation of stringent security protocols to prevent unauthorized access and data breaches. To combat this, we will explore encryption methods and secure storage solutions that comply with data protection regulations, thereby ensuring the confidentiality and security of student information while fostering trust within the user community.

Another significant challenge, with the necessity to verify the identity of HKU students when reviewing courses, in order to maintain the integrity and credibility of the platform. Utilizing HKU UID as a verification tool might seem like a feasible solution, yet it might entail complex integration with the University's existing systems. Moreover, too many verification steps might deter users due to the cumbersome process. Hence, one optimal solution involves a streamlined verification process using transcript or student card scans, which can offer a balance between security and user convenience, pending university approval and collaboration.

Furthermore, a potential challenge could be restricted access to the official HKU database, which might hinder the seamless integration of course details and student data into the platform. To circumvent this, the utilization of mock data can be considered as an initial step, paving the way for further negotiations with the university for eventual access to the necessary databases, and ensuring the accuracy & comprehensiveness of the data available on the platform. To ensure the platform remains up-to-date with evolving course offerings, course prerequisites, majors/minors, and many more. There is a necessity to develop admin APIs that allow for the easy addition of new courses in the future. This ensures the platform remains scalable and adaptable, catering to the dynamic nature of academic environments.

### **8.3 Machine Learning**

Data quality is directly related to the information gathered from course reviews and other sources and is one of the primary hurdles when it pertains to the training of our models. Incomplete or inconsistent data in reviews, for example, could result in unreliable insights for students. Noisy data, such as irrelevant or spam reviews, may affect the accuracy of course recommendations. To mitigate these issues, extensive data preprocessing and cleaning are essential to ensure that the platform provides students with trustworthy and valuable information for their course planning.

Data imbalance is also pertinent to our project, particularly in tasks like sentiment analysis and spam detection when dealing with user reviews. It refers to the scenario where one class of data (e.g., positive sentiment or non-spam reviews) significantly outweighs the other. In our case, it could mean an unequal distribution of positive and negative sentiment reviews or spam and non-spam reviews. Addressing this issue is crucial to ensure the effectiveness of our sentiment analysis and spam detection algorithms. We may need to employ techniques like resampling or adjusting class weights to balance the dataset, allowing our ML models to make accurate predictions for both classes.

Furthermore, algorithm selection is a key consideration in our project's machine learning components, particularly when designing algorithms for tasks like sentiment analysis, spam detection, and personalized course recommendations. Different algorithms have varying strengths and weaknesses depending on the type of data and problem at hand. To address this challenge effectively, we should engage in comprehensive experimentation with a variety of ML algorithms

that are well-suited to our specific tasks. By doing so, we can assess and compare their performance against each other and select the one that aligns best with our project's objectives.

Biases could also manifest as favoring certain types of courses or professors over others, potentially making this a disadvantage to particular groups of students. To address this challenge, we can implement techniques like re-sampling, which involves adjusting the training data to reduce the impact of biases. For instance, if there is an underrepresented group in the data, we can oversample or synthetically generate data points for that group to balance the dataset.

## 10. Project Management

The comprehensive development of our project has been strategically organized into several distinct domains to minimize overlapping responsibilities: Product Management, Development and Analytics, Machine Learning, and Quality. This thoughtful division of labor serves to optimize our project's efficiency and effectiveness. Each team member has been entrusted with specific roles and responsibilities within the aforementioned domains that are tailored to their unique expertise and strengths, ensuring effective and seamless collaboration.

The distribution of responsibilities among our team members is as follows:

<b>Name</b>	<b>Role(s)</b>
Md Abdullah Al Mahin	DevOps Engineer Back-end Developer
Richard Bryan	MLOps Developer Front-end Developer
Sabyasachi Purkayastha	MLOps Developer Back-end Developer
Yuen Yau Yoo	UI/UX Designer Front-end Developer QA Engineer

This structured allocation of roles and responsibilities will enable us to harness our collective potential and deliver a high-quality, well-rounded software engineering final year project.

## 11. Provisional Project Schedule

In order to effectively manage and execute our project, it is imperative that we establish a meticulously crafted schedule. This schedule serves as the roadmap that guides us through the various phases of development, ensuring that tasks are completed in a timely manner and project milestones are achieved with precision. In this section, we delve into the intricacies of our project schedule, highlighting key milestones, timelines, and the critical path that will lead us to the successful realization of our objectives. As such, the tentative schedule is presented below.

Date	Deliverable
1st October 2023	Project Plan Due
October	UI/UX Mock-ups, Database & Mock Data makeup, Data aggregation microservice, Relevance and sentiment analysis of reviews
November	Frontend Development, API set up, Data aggregation microservice, Spam detection of reviews
December	Frontend Development, API set up, data collection for ML recommendation
8-12th January 2024	First Presentation
21st January 2024	Preliminary Implementation & Detailed Interim Report
January	Frontend Development, AWS Infrastructure set up, ML recommendation & ML feature - related courses
February	Frontend Development, AWS Infrastructure set up & hosting, ML recommendation & UBA
March	Overall System Testing, AWS hosting
15-19th April 2024	Final Presentation

23rd April 2024	Finalized Tested Implementation & Final Report
26th April 2024	Project Exhibition

## 12. Conclusion

In conclusion, our COMP4801 Final Year Project is a substantial and ambitious initiative with the goal of transforming the course planning and registration process for students at the University of Hong Kong (HKU). This process seeks to simplify the complex task of choosing courses by introducing a user-focused course management platform enriched with advanced features. Our objective encompasses the creation of a centralized course repository that provides extensive course details, consisting of in-depth reviews and tailor-made suggestions. We strive to streamline scheduling choices, take into account graduation requirements, and academic aspirations, and offer refinement options for course recommendations. Furthermore, as mentioned previously, our platform will be integrated with spam detection algorithms, sentiment analysis, and user behavior tracking to enhance the quality and relevance of the information provided on the platform.

To pursue these goals, we have meticulously chosen a technological infrastructure consisting of a robust framework and database. Additionally, we placed significant emphasis on safeguarding data privacy, implementing authentication measures, and achieving seamless integration with HKU's system, all aimed at ensuring a smooth and user-friendly experience. Despite the expected challenges which include issues like data quality, potential model overfitting, and concerns related to data privacy, our project management strategy involves breaking down tasks into specialized areas, enabling each team member to make valuable contributions based on their expertise.

To summarize, our project aims to empower HKU students with a user-friendly platform driven by data to simplify their course planning and registration. We are committed to addressing the challenges ahead and delivering a valuable solution to the University community.