# Report 1625

**Name: Huynh Van Qui**

**Student ID: GCC210153**

**Class: GCC210153**

* Topic name: **Application of Big Data & AI in Construction and Exploitation of National Database of Real Estate Prices**

## Assignment 1

* **Topic**: My team has been tasked by the government to undertake a significant project that merges big data processing with AI. This project aims to manage large databases from the Ministry of Finance and the Ministry of Natural Resources and Environment for real estate valuation. AI is utilized for project management and operation, integrated into a mobile app for easy monitoring from various locations. AI also processes information, provides solutions, and supports different customers. Big data is crucial for quickly extracting and storing vast data from the mentioned ministries.
* **My topic can solve problems like:**
  + The project can collect data from databases of the Ministry of Finance and the Ministry of Natural Resources and Environment and then filter these data to select the data needed.
  + The system can estimate the price of real estate for users. This price will almost be well-price in the market.
  + User data have to be protected. It is hard to hack the sensitive data of users and the system.
  + The system can depend on the behavior of users to suggest real estate.
  + Almost all real estate is included in the system, so users can search it easily.
  + The system can analyze and process large data to respond to user requests quickly.
  + The system will provide the owner's contact information to people who want to buy real estate, but only if the owner agrees to provide the information. This will ensure the security of user information in the system.
* **My plan**
  + **Scope:**
    - **Acceptance criteria in my project:**

The system must update data quickly and accurately from government databases, such as the Ministry of Finance and the Ministry of Natural Resources and Environment, which are responsible for collecting and managing data on land, property, tax, and environment. In addition, the system must ensure information security by protecting the personal and financial data of users and owners from unauthorized access, modification, or leakage. Moreover, the system should comply with the relevant laws and regulations on data protection, such as the Law on Cybersecurity and the Law on Personal Data Protection. The system should message the user who pays attention to the real estate they need as fast as the system can, by using AI techniques such as natural language processing, sentiment analysis, and recommendation systems. The system should understand the user’s query, analyze the user’s preferences and emotions, and suggest the most suitable properties based on the user’s needs and budget. The system is required to maintain its smoothness and stability even when used continuously over an extended period, by using big data technologies such as cloud computing, distributed processing, and data engineering. The system should be able to handle large volumes of data, process complex calculations, and deliver fast and reliable results.

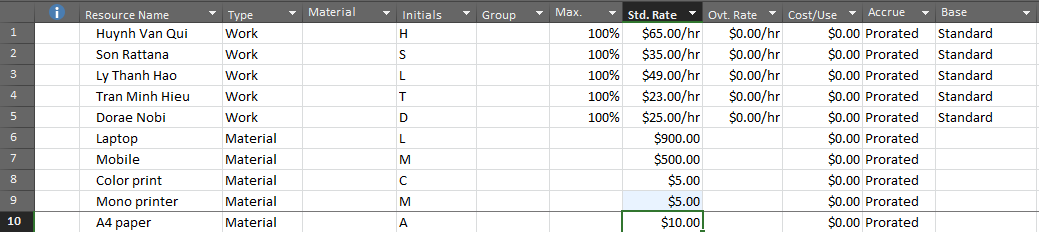
* **Exclusions of the project:**

The project will not be responsible for the costs of maintaining, changing, or upgrading the system. The project will only deliver the system as a product, but not as a service. The users or owners of the system will have to bear the costs of operating, maintaining, changing, or upgrading the system according to their needs and preferences. The project does not pertain to the online buying and selling of real estate. The project will only provide information on the real estate prices, but not mediate the transactions between buyers and sellers. The users or owners of the system will have to use other platforms or channels to buy or sell real estate properties.

* **Time:**

The project is expected that the project will take 171 days to complete.

* Date start: 20/9/2023
* Date end: 16/5/2024
* **Milestone**
* General survey: 13/10/2023.
* Plan for the project:3/11/2023.
* Build a complete system: 10/4/2024.
* Project handover: 16/5/2024.
* **Communication**: The project implementation process will have 2 meetings including Project update meetings and Progress project.
* **Risk:** While analyzing the possible threats to the project, I found 5 aspects that could pose a risk:
* There’s a possibility that technology might not live up to performance standards, experience issues with compatibility, or confront unexpected technical hurdles.
* Going over the budget due to unexpected costs during the execution phase.
* Incompatibility with mobile operating systems can result in operational interruptions, technical glitches, and inadequate scalability, which can negatively affect system performance and customer satisfaction.
* Breaches of data privacy, unauthorized access, and cyber-attacks can result in the compromise of sensitive system and user information.
* Risk of user data leakage.
* **Costs for personnel and materials:**

****

To complete the project in 171 days, the total amount of money needed is 94,392 USD

* **Paper:**

My project was based on the two articles I studied, which improved the project Big Data and AI to create and use a national database of real estate prices. I employed the HPM, a common method to assess the value of real estate properties according to their features and location, and used various sources of big data, such as the internet and remote sensing, to build and exploit the database. I also applied big data technology to enhance the HPM, such as data pre-processing, spatial modelling, GIS spatial analysis, and machine learning methods. These techniques helped me overcome the difficulties of real estate big data, such as high dimensionality, heterogeneity, sparsity, and noise. Moreover, I explored the connection between real estate price and mobility patterns, using CDR as a new source of big data. CDR can offer information on the activity and entropy of dwellers and workers in different regions, which can influence the demand and supply of real estate properties. Furthermore, I used AI to model real estate price using CDR and hybrid machine learning approaches, such as MLP trained with PSO. AI can help the project increase the accuracy and efficiency of real estate price prediction, as well as generate recommendations for improving the real estate market conditions.

## Assignment 2

* Logbook:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Project Logbook** | | | | | | |
| Begin date: 20/9/2023 | | | | End date: 16/5/2023 | | |
| Scheduled | | Actual | | Task action list | Phase name | Problems and Changes |
| Begin | End | Begin | End |
| 20/9/2023 | 13/10/2023 | 20/9/2023 | 13/10/2023 | * Define project objectives * Define project scope * Define User requirement * General Survey | Initiation Phase | This phase has some issues related to collecting the requirements of the customer. |
| 13/10/2023 | 3/11/2023 | 13/10/2023 | 4/11/2023 | * Define Project Plan * Define Risk planning * Define Resource * Define Financial * Plan for the project | Planning Phase | This phase had planning problems so it took 1 more day than expected. |
| 3/11/2023 | 5/12/2023 | 4/11/2023 | 6/12/2023 | * Find and collect pertinent literature * Develop research from pertinent literature * Identify and collect Qualitative Research * Identify and collect Quantitative Research * Survey research * Data analysis * Result analysis * Collect information and data | Research & Analysis | This phase has some difficulty in finding the study about AI and Big Data related to estimating real estate, but my teammates are so good thus this phase is done on the timeline. |
| 5/12/2023 | 10/4/2024 | 6/12/2023 | 10/4/2024 | * Design Phase: * Design graphic user interface website and mobile * Design Database * Design system architecture * Implementation Phase: * Font-end Development * Collect data from other databases * Data synthesis * Database development * Develop construction of real estate price index by AI. * Testing * Deployment Phase * Data migration * System deployment * Build a complete system | Executing Phase | To make up for the time spent in the previous stages, my team had to put in more effort at this stage. Therefore, the phase is completed sooner than 1 day. |
| 10/4/2024 | 16/5/2024 | 10/4/2024 | 16/5/2024 | * Write Final Report * Write User Document * Prepare for Presentations * Send the Final report to all parties involved * Project evaluation * Conduct Project Review and Identify Lessons learned * End the project and save data * Project handover | Closing Phase | We had some issues while we wrote the user document, but the team discussed to solve it. Thus, this phase is also done on time. |

* **Advantages and disadvantages of my topic**
  + Advantages:
    - Big Data and AI can help the project construct and exploit a national database of real estate prices, using various sources of data, such as the internet, and remote sensing.
    - Big Data and AI can help the project optimize the hedonic price model (HPM), a widely used method to estimate the value of real estate properties based on their characteristics and location.
    - Big Data and AI can help to improve the accuracy and efficiency of real estate price prediction, as well as generate recommendations for improving the real estate market conditions.
    - Big Data and AI can help the project gain insights into the relationship between real estate price and mobility patterns, using call detail records (CDR) as a novel source of big data.
  + Disadvantages:
    - Big Data and AI can pose ethical challenges, such as privacy, discrimination, social cooling, big data divide, and social sorting.
    - Big Data and AI can be costly to implement and maintain, requiring skilled personnel, advanced hardware, and software.
    - Big Data and AI can lack emotion and creativity, which are essential for human interactions and problem-solving.