# MSc AI Dissertation Project Proposal Form

|  |  |  |  |
| --- | --- | --- | --- |
| Student number | 202249387 | Date | 20 June 2024 |
| Student name | Timothy SB Hallworth | | |
| Supervisor name | PA Robinson | | |
| Project title: | An analysis of the Property Market in Hong Kong using Machine Learning | | |
| Project description – background, aims, objectives, research question(s), brief reflection on novelty/originality | | | |
| The Hong Kong property market is in a depression, and I would like to use machine learning to analyse and predict market trends.  The aim will be to use machine learning to identify the key factors affecting property prices and demand and predict future trends.  The objectives are to collect and preprocess property data, implement machine learning models for analysis, evaluate model performance, to identify significant market trends and finally to predict future property prices and demand.  The research questions I would like to address are:  1. What factors influence property prices in Hong Kong?  2. How accurately can machine learning predict future prices and demand?  3. What trends can be identified from the data?  This project uniquely applies advanced machine learning to the Hong Kong property market, offering deeper insights and more accurate predictions than traditional methods. | | | |
| Methodology – rationale, data selection and collection, recruitment of participants, analytical process | | | |
| I would employ machine learning techniques to uncover patterns in the Hong Kong property market that traditional methods miss.  I will collect historical property transaction data from reliable sources such as Century 21, government databases, and other real estate websites. Key features will include property prices, location, size, and other relevant factors.  I do consider there is a need to recruitment human participants as all the data will be sourced from public and commercial databases.  The analytical process would be (i)data preprocessing (which would include cleaning and normalising the data and encoding categorical variables) (ii) Feature selection using techniques like PCA and Lasso Regression to identify critical features; (ii) model implementation by applying a machine learning models such as a Deep Neural Network; (iv) model evaluation using cross-validation and metrics like RMSE, MAE, and R² to assess model performance; and (v) Using a trend analysis to identify and predict market trends based on the model outputs. | | | |
| Theoretical framework and intended tools/technologies | | | |
| The theoretical framework would use deep neural networks to predict property prices.  The intended tools and technologies would be to use Python for coding, to handle the data using Pandas and NumPy, to visualise the data using Matplotlib and Seaborn and to use the machine learning libraries of Scikit-Learn and TensorFlow/Keras for deep learning. | | | |
| Key literature identified for literature review (include references) | | | |
| To be determined, I would need to consider the literature on machine learning in real estate, the feature section of real estate, data processing real estate information, and a review of real estate analysis. | | | |
| Industrial partner (if appropriate) | | | |
| Not applicable. | | | |
| Project management | | | |
| Table: Project timeline and key outputs (expand table as needed)   |  |  | | --- | --- | | Week | Activity | | 1-3 | Project Setup and Data Collection  - Define project scope and objectives.  - Collect historical property transaction data.  - Set up the development environment | | 4 | Data Preprocessing  - Clean and normalise data.  - Encode categorical variables.  - Handle missing values. | | 5 | Feature Selection  - Apply and consider regression for feature selection. | | 4 | Model Implementation  - Implement Linear Regression and Decision Trees using Scikit-Learn.  - Build and train neural networks with TensorFlow/Keras. | | 5 | Model Evaluation and Trend Analysis  - Evaluate models using cross-validation, RMSE, MAE, and R².  - Analyse significant market trends and patterns. | | 6 to 10 | Documentation and Presentation  - Compile results into a comprehensive report.  - Create visualisations with Matplotlib and Seaborn.  - Prepare and deliver the project presentation. | | | | |
| Resources and Research Data Management Plan (Describe the resources required and the data you expect to acquire or generate during this research project, how you will manage, describe, analyse, and store the data and what mechanisms you will use to share and preserve your data.) | | | |
| I have the resources for hardware and software. I need to consider the data sources which may be the biggest problem however, I believe there will be a lot available.  I would site the data on my computer and in the cloud. I already have used analysis tools such as Pandas, NumPy, PCA, Lasso Regression, Scikit-Learn, TensorFlow/Keras.  I do not plan to share it with anyone except the academic staff at the University of Hull. | | | |
| Planned outputs/publications/research datasets/impact/dissemination | | | |
| At this stage the outputs would be a research paper and a power point presentation. | | | |
|  | | | |

|  |  |
| --- | --- |
| If successful, I undertake to carry out the research according to the University’s Ethics code of Practice. I realise that I will not proceed into data collection without an ethical approval in place  (Applicant’s signature required) |  |
| Date and signature of Supervisor approval |  |