Why do you want to research on this topic?

I choose to do research on this topic because there is a substantial need of empowering potential customers with personalized recommendations in real-estate sector using Machine Learning and Computer Vision. As we know that generating appropriate recommendations for the user by analyzing available data has grown vastly over the past few years e.g. how stocks of a company will go in the near future. But in the real-estate sector, the researches available are not sufficient.

What kind of data will you be utilizing?

We will have two sets of data, text-based data and image based. The text-based data will contain customers search history, property view history and past investments history at Mason Verdi. The image-based data will contain the images of the buy-to-let investment properties of Mason Verdi.

What will be the project objectives?

We will utilize Machine Learning and Computer Vision to offer personalization for the buy-to-let investments for the customers on the basis of various sets of customers search history, property view history, past investments history and the images of the buy-to-let investment properties at Mason Verdi.

What is your proposed Methodology?

We will be using similarity learning from user behavior which is one of the fundamental concepts of machine learning for building recommendation systems. We can define a distance or similarity measure, e.g. cosine similarity of raw features of the buy-to-let investment properties of Mason Verdi. A potential customer viewing two properties listings in quick succession will offer some evidence that they are similar to each other. When aggregated over the behavior of lots of users, anonymized listing browsing data will offer an informative similarity signal we can use to train our recommendation system model.

What tools and technologies will you be using?

We will use Machine Learning algorithms, Computer Vision, heuristics, a database to store and retrieve all sets of data and a cloud server for testing and deployment of the recommendation system.

What will be your project flow?

The first step is thorough literature research. Once that is done, the initial approach we would take for similarity learning could be to train a binary classifier to classify a pair of properties as similar or not similar. Properties listing viewed in quick succession will be more similar than listings viewed far apart, because they will likely express the same user interest. However, we would not know the magnitude of that similarity at this point. So, we will frame this as a learning-to-rank (LTR) problem with the goal of presenting users with the most similar properties, ranked in descending order of similarity. For the LTR problem, we will use pairwise and listwise loss functions. For optimization of the developed similarity learning, we will utilize heuristics to help safeguard the model predictions by providing guardrails against anomalous outcomes e.g. we can have heuristics on basic features of the properties such as bedrooms, bathrooms, and price to enforce an absolute delta between the anchors and the candidates on those features. Finally, we can perform Ranker-Based Ensembling to make a one cohesive model.

What will be the impact on the world of science when this PhD project will be completed?

The impact on the data science world by the time this PhD project will be competed will not only the knowledge contribution regarding learning similarity from user behavior using Machine Learning model as a ranking problem but it will go beyond that and will contribute knowledge regarding several considerations to optimize for similarity including heuristics and ranker-based ensembling and extensibility to make efficient design decisions to combine similarity learning and ranker-based ensembling and extensibility into one cohesive model.

What difficulties would you expect to encounter during this project?

The challenges for the project would be gathering relevant data for training and getting accurate ranking from the trained model.

How would you overcome these challenges?

We would overcome the data challenge by corresponding with Mason Verdi to check if they have the required data. If not, we will generate the data in Python or MATLAB. To get accurate ranking, we would perform content based filtering in which we would create a clustering algorithm to create interest clusters from a user’s view history.