**Smart transportation planning: Data, models, and algorithms**

Over the last 20 years, demographic and economic changes have pushed cities and urban centers to become the main habitat of humankind.

Cities are drivers for the countries’ socio-economic prosperity. On the other this booming in urbanization and cities growth have exerted enormous pressure on transportation systems around the world which in turn transferred part of the pressure to the environment. It is, therefore, more critical than ever that public and private parties commit themselves to realize the goal of sustainable urbanization.

The journal is presenting on high level what different machine learning models can be used to help mitigate the transportation challenge. Multiple models were mentioned including Clustering, ARIMA, KNN, KD, Holt-Winters’ exponential smoothing model, Random walk model and deep learning. The models were deployed mainly to forecast traffic flows and predict some associated events that could impact traffic situation, such incidents like weather condition, rush hours, and accidents. In addition to forecasting real estate prices.

The main objective of selecting this paper was to identify what common models being used and how they are applied to perform predictions/forecasting and classification. In our case we have data related to construction and population so this paper will help us perform similar predictive models to forecast urbanization growth and predict emerging construction hotspots. Moreover, we could predict real estate prices in case we managed to find related prices information

However, the challenge will be in finding enough relevant data that can be used to come up with forecasts with acceptable accuracy. Relying on multiple data sources could be one way for overcoming this challenge. Another challenge would be the complexity of the models. Therefore, exploring existing machine learning platform like Google cloud and Microsoft Azure could save some of the hassle and time.

Link to the journal:

<https://www.sciencedirect.com/science/article/pii/S2666691X20300142>

**HomeSeeker/ A visual analytics system of real estate data**

Finding suitable home or property is becoming crucial for many of us. Moreover, advancements in technology are shaping the way real estate agents and homeowners navigate the home selling and buying process. However according to the author, despite the convenience that existing systems are providing, they still lack many of the important features that every home seeker requires. Some external information like neighborhood facilities such as schools, stations, supermarkets, and hospitals are not available online within the majority of the real estate portals.

The main idea of the paper is to build a tool that provides home-seeker a convenient way to look for his dream house by bringing a 360 view of the property itself and the neighborhood. In addition to providing real estate market update all in a visually appealing way.

The tool also integrates multiple sources together including basic information like prices, address, number of rooms, in addition to variety of profiles such as education, transportation, facility, regional and environmental. All data are aligned and presented in multiple visualization techniques with a high-level guidance on the best use of every one of them.

The variety of visualization included in the paper had provided us a complete view on different visualization designs and a recommendation about their uses some of those designs included, **Choropleth map, multiple variety of Dot maps, Glyphs on map, those are for geo-related visualization. On the historical visualization the study mentioned, Multiple line graphs Stream graphs, histogram, in addition to other general visualizations like word cloud, Parallel coordinates, Colored Boolean table and Geo-coded scatterplot.**

The paper will be extremely helpful to us to identify and explore different visualizations designs for different presentations. Moreover, it will give us high level guidance on deciding what design to be used with which data to better visualize and summarize the insights or information.

The potential challenge is how to integrate and stream line multi data sources into one aligned dataset. Adding to this the level of complexity of some of the graphs in both visualization and interactivity. To overcome this, we are planning to utilize third party visualization tools like Tableau and Power BI.

Link to the journal:

<https://www.sciencedirect.com/science/article/pii/S1045926X17301246>

**Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model**

Urban growth is a worldwide phenomenon and the rate of urbanization is very fast especially in developing countries like India. It is mainly driven by rapidly increasing population, and unorganized expansion. Given the rapid urban growth and importance of its long-term effect, it is becoming increasingly important to monitor and analyze the urban land cover change, as well as to adopt appropriate sustainable land use plans. This paper presents the use of machine learning algorithms to monitor and predict Mumbai city expansion. The study was able to identify the years where urbanization change peaked and the most affected areas like open lands and croplands. In addition to identifying the impact of some of the parameters like regulatory intervention on the urban growth dynamics.

According to the authors, they took a unique approach in selecting the models for their prediction. They basically integrated two models, the Markov chain which is intended to determine the actual amount of change between land use categories non-spatially and the Cellular Automata that is used in spatial allocation and location of change. They claimed that each model has its own strength and limitations and by integrating both models helped in overcoming some of those limitations

The study collected data from multiple sources including satellite land images, population data, Digital elevation model from NASA and lastly transportation network data. They started by classifying satellite images then computing the transition probability maps on the basis of auxiliary data, based on MCE (multi-criteria evaluation). They also added land use map to predict future urban growth for 2020 and 2030 using Markov chain and Cellular automata.

In our project we could use such blend of models to help us predict or forecast the construction hot areas taking in mind proper urbanization planning. Another area where we could explore this approach is in forecasting urbanization growth as a whole.

Our challenge in adopting similar approach would be the availability of data such as satellite images. However, we could rely on population data and try to find other relevant public data to add it to the dataset we have.

Link to the journal:

<https://www.sciencedirect.com/science/article/pii/S0143622813000362>