Optimization of Single Mode Trip: Technique-Based Recommendation System of Machine Learning

Chaimae Attar and Najima Daoudi

October 3, 2021

ATTAR CHAIMAE, DAOUDI NAJIMA

National Graduate School of Computer Science and Systems Analysis, Rabat/Morocco

Chaimaeattar96@gmail.com

Abstract. Smart city is a recent topic, but it is developing very rapidly, as it is seen as a winning strategy to deal with some serious urban problems such as traffic, pollution, energy consumption, waste treatment.

In this context, the present brief aims to develop a system for recommending the best routes for passengers according to their departure and arrival addresses .To meet this objective, it is necessary to carry out an analysis in order to define the different tools and methods to be used. In addition, after the identification of user behavior needs. We carried out a design adequate to our recommendation system and well detailed for each module. Then we compared the different models RandomForest, artificial neural networks, KNN Basic, KNN Means, KNN ZScore,SVD. Finally, we found that the two models RandomForest and artificial neural networks are the mostefficient compared to the other models, with an accuracy of 0.97 for the first one and 0.90 for the second one.

Keywords: Recommendation system, deep learning, artificial neural networks, Machine Learning,

intelligent mobility

1 Introduction

The growth in the population of cities and the increase in the number of road users is a considerable source of road safety problems caused by accidents and incidents caused by deteriorating traffic conditions. As a result, the human and financial damage increases in addition to the time lost in traffic jams and the inconvenience to road users and the lack of traffic flow. In order to face this road safety problem, There are several projects to be undertaken, including the improvement of traffic conditions. Indeed, the implementation of an intelligent transportation system capable of offering optimized modes of transportation and routes to users could help overcome this problematic situation.

Our research focuses on road safety and aims to consolidate the fluidity of traffic between cities and thus improve passenger satisfaction. In this paper, we deal particularly with the issue of smart mobility in order to guide users towards a personalized mobility service, adequate and adapted to their needs. Smart mobility is a component of the smart city that has revolutionized the way we think about travel to be safe, fluid and efficient. To do so, it would be wise to accompany road users and guide them towards solutions that are appropriate and well adapted to their needs.

The massive data available to users and their preferences, as well as unprecedented storage capacities, have favored the proliferation of intelligent services, especially smart cities. Indeed, Artificial Intelligence (AI) uses sensors and actuators such as video cameras, environment detection, traffic detection, smart meters, vehicles, as well as smart phones. These large and fast data can used to extract exploitable information where AI can shine and improve the quality of life and urban services.

For the rest of our paper, it will organized as follows: In the second section, we discuss the state of the art of the concept of 'Smart Mobility' in order to understand the context of the research and to situate our problem in it. In the third section, we will focus on AI techniques (Machine Learning, more precisely Deep Learning used based on the papers). In the fourth section, we will propose a recommendation system for the optimization of single-modal trips, i.e. the use of a single means of transport for the trip from point A to point B. In the last section, we will evaluate the two models in order to choose the most adequate recommendation system to guarantee a better safer and fluid mobility.

2 State of the art

In order to optimize single-mode routes, we have chosen to use a recommendation system based on AI techniques. Therefore, we will present all the terminologies and concepts. We will start with the definition of intelligent mobility. Thus, we will present a topological overview of recommendation systems and their techniques.

2.1 Smart Mobility

Smart Mobility refers to the application of new technologies in the field of transportation; it is characterize by the set of means aimed at protecting the environment, fighting against the climate change. In addition, the road safety and transportation management field are among the fields of application of intelligent mobility by encouraging ways of alternative transportation such as bicycles and carpooling platforms. Indeed, Smart mobility has for objective:

- Operate the entire transportation network;
- Ensure the safety of travel;
- Add value to existing mobility services;
- Ensure the accessibility of services;
- Reduce the environmental impact of transportation.

2.2 Recommender systems and their techniques

The definition of a recommender system has evolved over time as the field has progressed. That the field has undergone. A recommender system helps users make choices in a domain where they have some information to sort and evaluate possible alternatives (Resnick & Varian, 1997; Shardanand & Maes, 1995). A second, more popular and general definition is that of Robin Burke, who defined recommender systems as systems capable of providing targeted recommendations by analyzing user's preferences on items in order to suggest and guide them to choose the products or services that best meet their needs. Indeed, recommendation systems aim at providing users with relevant resources according to their relevant resources according to their preferences and therefore they will allow reducing the search time (Burke 2002).

3 Machine and Deep Learning algorithms for user route recommendation

Nowadays, the most famous recommendation systems in the world such as Netflix, Amazon use Machine and Deep Learning techniques to improve the efficiency of their systems. Indeed, we have chosen to evaluate the Random Forest, Support Vector Machine "SVM" and Artificial Neural Network "ANN" algorithms for the content-based approach and k nearest neighbors "KNN" and the matrix factorization for the collaborative filtering-based approach that we will detail in the next sections.

3.1 Random Forest

Random Forest is composed of several individual decision trees. Each tree makes a class prediction and the class with the most votes becomes the prediction of our model. In effect, a decision tree is viewed as a set of yes/no questions asked, for the data in order to find the predicted class, it tries to form nodes containing a high proportion of samples of a class by looking for values that divide the data well into classes. One of the works that showed the effectiveness of the Random Forest algorithm for recommendation based on the rating is a work done to predict the score of movies in the context of online sales; this system aims to personalize the recommendation for each user according to their preferences and their interests. The author evaluated several algorithms: Euclidean distance, K-Means and Random Forest, of which Radom Forest showed 96% accuracy (Jayashree Nair 2016).

3.2 Support Vector Machine « SVM »

Data scientists often use SVMs for classification tasks, and they have worked well in a variety of have worked well in a variety of problems. Indeed, SVM classifies data by determining the optimal hyper plane that separates observations based on their class labels. Thus, the central concept is to account for classes separable by both linear and non-linear class boundaries.

Moreover, non-linear class boundaries. Vapnick developed the algorithm in 1993. The author Gilles Lebrun in 2006 used the SVM algorithm to develop an application for image processing and analysis whose goal is to produce a powerful decision process and take into account the fact that the learning time is exploitable compared to the size of the data.

3.3 Artificial Neural Network

Artificial neural networks are a non-linear statistical data model that used to model a complex relationship between input and output for pattern matching. Indeed, ANNs have often been implemented in many scientific fields for prediction purposes (Williams, J.; Li, Y 2008). In road safety research, neural networks have used to study accidents based on driver, vehicle, road and road condition characteristics (Abdelwahab, H.; Abdel Aty 2001).

In 2017, researchers proposed a new methodological approach to road safety risk, which is a two-stage framework consisting of Data. Envelopment Analysis (DEA) in combination with artificial neural networks (ANN). In the first phase, the risk level of the studied road segments calculated by applying DEA, and the high-risk segments identified. Then, the ANN technique adopted in the second phase, which seems to be a valuable analytical tool for risk prediction. Indeed, the practical application of the DEA-ANN approach in the geographic information system (GIS) environment was an effective approach to road safety risk.

3.4 K-Nearest Neighbors

This algorithm used for classification and regression problems in Machine Learning. KNN is also non-parametric. Indeed, the K-NN algorithm relies on the entire dataset to perform the prediction. In fact, for each new observation that we want to predict, the algorithm will look for the K instances of the dataset that are closest to our observation. Then, the algorithm will use the output variables of the neighbors to compute the value of the observation we want to predict. Several recommendation systems based on collaborative filtering inspired by this algorithm in order to make the recommendation. One of the works that used K-NN for recommendation are the two authors Adeniyi and Yongquan in 2014, in order to rank users based on their current click stream, and recommending a personalized browsing option that satisfies the needs of the active user at a given time. All this done by applying the web data collection technique to extract the knowledge needed to provide real-time recommendation services on the site.

3.5 Matrix Factoring

Matrix factorization allows to process large amounts of data in a fast and accurate way. This pure notation based approach popularized with the Netflix award. It uses algorithms that are not complex to implement and efficiently. The principle of matrix factorization is to decompose a matrix into several other matrices. The original matrix is the product of these matrices between them. Indeed, the decomposition of a matrix designates both the products and the users by vectors. Then, the recommendation is the product of the products and the user items. Indeed, "Netflix Price Challenge" demonstrated that matrix factorization models are superior to classical nearest neighbor techniques for generating product recommendations, allowing for the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels (Sachin Prabhu, ThandapanI 2019).

4 Recommendation system design

Smart mobility consists in recommending to passengers and drivers the offers that meet their needs; its main objective was the implementation of a recommendation system for the optimization of monomodal trips using Machine and Deep Learning techniques. To do so, we deconstructed our main objective into three operational objectives:

- Inform and guide users on the right mobility usage that fits their needs;
- To provide users with several means of transportation: cars, buses, CTM...;
- Accompany the person towards a collaborative mobility.

Our goal is to offer a solution for the entire population, whether passengers or drivers. To do this, we propose a global architecture illustrated in figure 1.

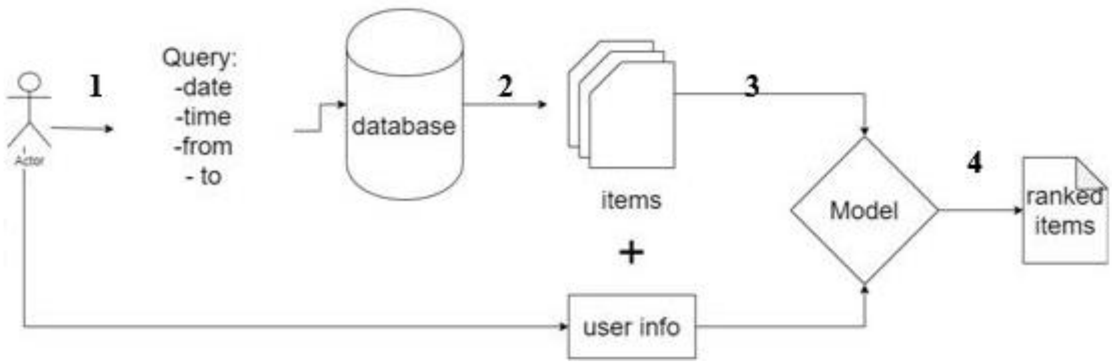

Figure 1: The architecture of the recommendation system

1. The first component of our architecture is the user. The user launches a request specifying the source, the destination and the date.

2. The system searches for available routes in the platform.

3. The system applies the model to the available trips by integrating the user's identifier to predict the rating of the trips.

4. The system sorts the obtained results from the highest to the lowest scores.

We then designed the technical aspect of our solution, illustrated in Figure 2, in order to detail the process of extraction, processing, implementation and recommendation of trips. Indeed, after the user enters the query with the source, destination and date, the system searches for the routes that meet the requirements. Then, we pass the obtained items in the pre-processing module where we transform the categorical data (day of the week, means of transport, route) into binary data and we perform a normalization, then we apply the model and after several adjustments, we obtain the prediction of the scores to sort the trips with the highest scores to the least scored.

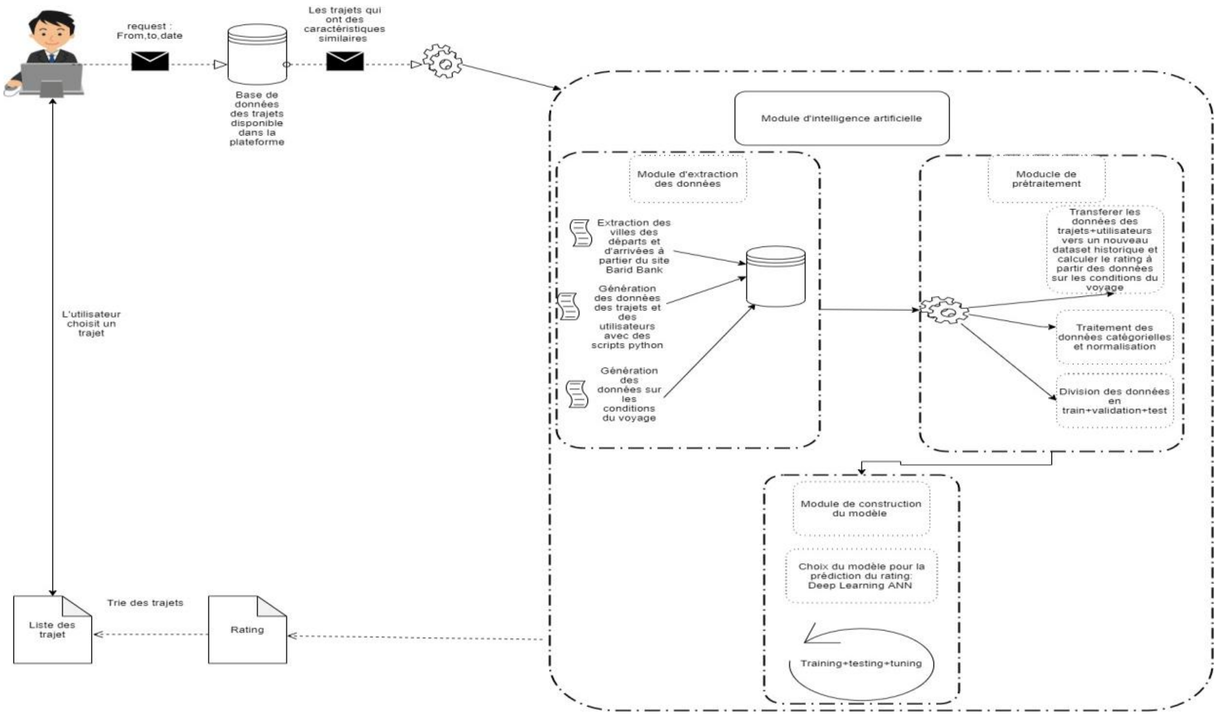

Figure 2: The technical architecture of the best route recommendation system for passengers

5 Implementation of the recommender system

In this section, we will explain the steps required for the implementation of a recommender system for single-mode route optimization using Machine and Deep Learning techniques.

5.1 Needs analysis

In order to understand the needs and preferences of the users and in the absence of real data, we conducted a questionnaire containing questions about the passenger’s preferences, the criteria for choosing the means of transportation...

After this study, we found that 92% are passengers and 7% are drivers. In addition, 52.3% of the users choose the fastest means of transport, and just 29.9% who chose the cheapest means of transport. Figure 3 shows the pie chart of the results obtained:

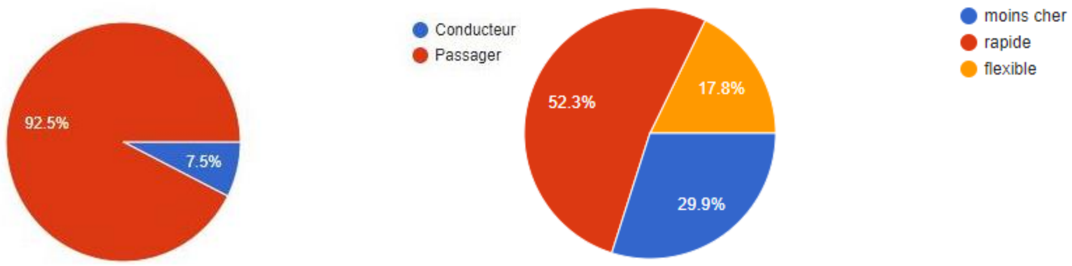

Figure 3: Survey results

Therefore, we can conclude that the users prefer the routes with less traffic jam and the fastest means of transportation like Tram and train.

Indeed, most of the users are between 20 and 31 years old, which proves that young people are the most active popularity, so we found that the most frequented region is RABAT with a percentage of 37.5%, because a very important number of road users knows it. Figure 4 shows the age and the cities frequented by the passengers who filled in the form:

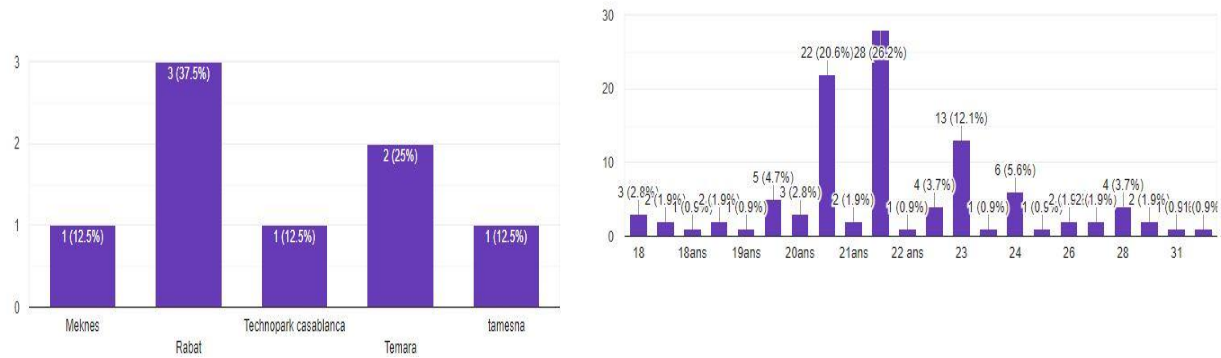

Figure 4: The presentation of ages and cities frequented by passengers

5.2 Technical environment

We used the python programming language in order to develop our solution because it is the most widely used languages in the data science field. In fact, Kaggle[1] conducted a global survey in October 2018 among 23859 data professionals, and it found that 83% of data professionals used Python, as its automatic memory management, its adaptation to several computing platforms without any hassle, knows it. This language also known for its performance in all Machine and Deep Learning problems, because it offers a wide choice of libraries to meet all the needs collected.

We worked in the Google Colab environment in order to reduce the execution time and avoid the problem of memory saturation and data loss in case of an incident.

5.3 Elaboration of the data sets

It should note that one of the constraints we found was the lack of data. Therefore, we had to generate our own datasets respecting the business rules in order not to distort the analysis. We generated three datasets, the first one is the trip data: we divided the trip data into two categories: the first one is the trip data. We started by generating the geographic data using the Google Maps API, then we generated the distances and times also with the Google Maps API. Then we generated the dates and times with the Faker library. It is a library that allows you to generate random data according to the type you specif. and then we added the durations to the generated dates in order to calculate the arrival dates. Concerning the durations, we calculated the duration for each means of transport. Then we calculated the price for each means of transportation by multiplying the distances by a coefficient. The second category concerns the travel conditions. In order to evaluate the quality of a trip we

[1] Kaggle. Available at: https://www.kaggle.com/

generated data on traffic conditions such as traffic, safety, accidents... Finally, we computed trip scores based on the conditions.

The second dataset we generated is the user dataset, so we used the Mockaroo platform for the user data generation. In addition, the last data set is the historical dataset in which we assigned each user 200 trips.

5.4 Implementation of the prediction process

After the data generation and before detailing the prediction process, it is necessary to present the approaches of a recommendation system as well as the algorithms that we used. Indeed, for the content-based approach where the user does not intervene, we used the two Random Forest algorithms and the artificial neural networks. For the second approach based on collaborative filtering and which consists in recommending an item to a user based on similar users, we used K-NN and matrix factorization.

One of the essential steps in a Machine Learning project is a data exploration with visualizations to understand the structure and distribution of the data. After several displays, we noticed that most of the trips taken by users have a score between two and four. In addition, users use cars more than other means of transportation. In addition, passengers use highways more than the national road. Figure 6 shows a visualization of the explored data:

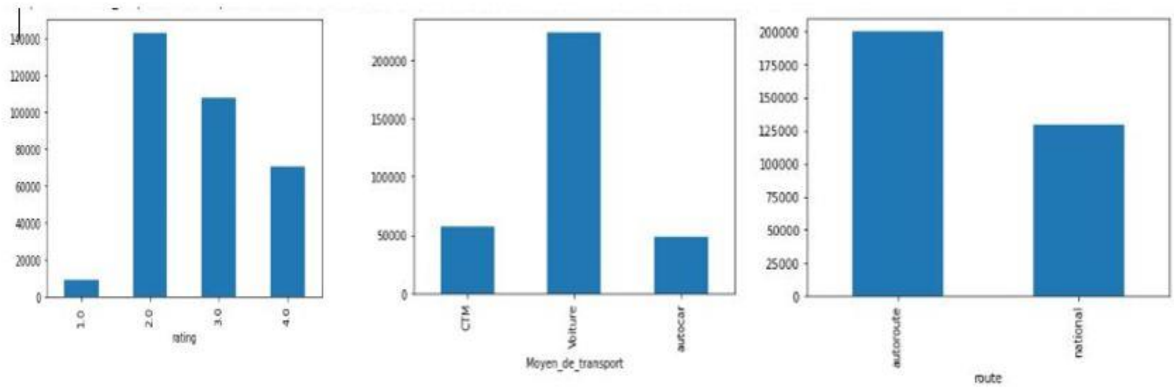

Figure 6: Data Exploration

5.4.1 Content-based approach

Content-based recommender systems are systems based on a user's evaluation of a set of products. The goal is to understand the reason that motivates a customer to judge the relevance of a given item. This approach based on the comparison between a product and the user's profile, the aim of which is to propose products that will be similar to those that the user has liked or purchased in the past. For this reason, it is necessary to describe the set of products through certain characteristics (for example : color, price, etc.). In this way, it will be possible to determine which characteristics are important in the customer's past purchases, and to recommend products that have the same properties. To address the content-based approach, whose main objective is to recommend optimal routes to users based on the history of the routes they have made, we chose to evaluate the two algorithms Random Forest and ANN.

5.4.1.1 Random Forest based model

In our project, we used the trip dataset with 21 features (departure city, arrival city, duration, distance, date, price, means of transport, route ...) and 688719 data points (samples) divided into 5 different labels (score from 0 to 4) of the trips made by the users. This is a multi-class classification problem, which means that the data are not linearly separable. For this reason, we used Random Forest for the classification of the trips. Indeed, we used 10 estimators in a first step and 70% of the training data and 30% of the test data but we found an overfitting problem. To solve the problem we applied a cross validation with five splits to find an accuracy of 97%.

After implementing the model, we displayed our confusion matrix, which is a performance measure for a machine learning classification problem where the output can be two or more classes. We can conclude that our model predicts class 0 with 90% accuracy and the probability of predicting it as class 1 is 9.84%. For class 1 or rating equal to 1, our model predicts it with 88% accuracy and 1%, as a class 0 is rating equal to zero, for the other classes, the model predicts them with 100% accuracy. Moreover, to understand the features that affect more on our model we have displayed 'features importance'. Figure 7 shows the confusion matrix and the importance of each feature in our model implementation.

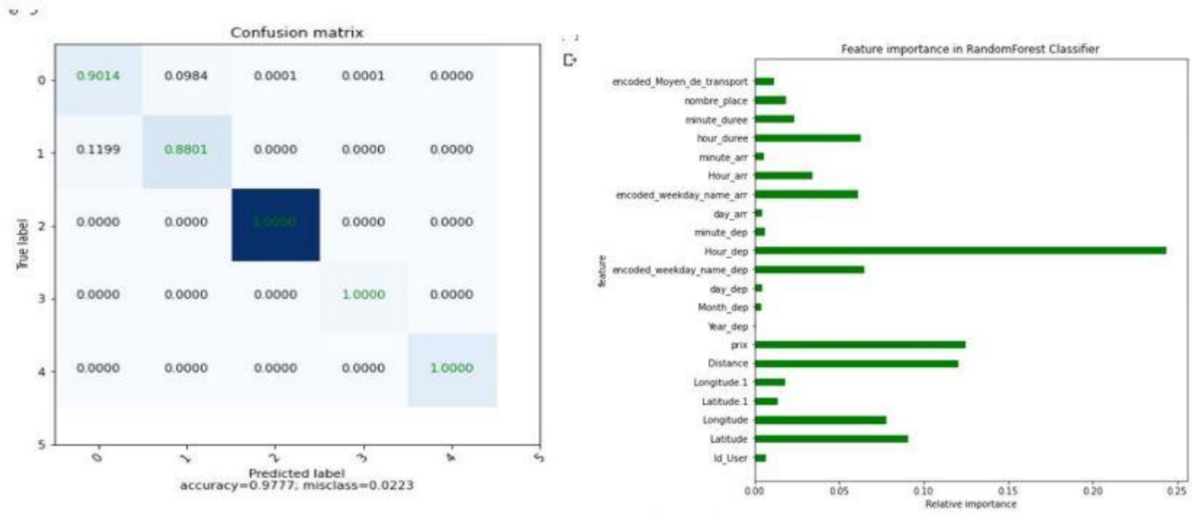

Figure 7: The confusion matrix and the importance of features

The features that have the greatest impact on our models are departure time, price, distance, day of the week. It is true that these are the features that measure the quality of a trip; that is, an optimal trip is one with the best price-distance combination and on days and hours when we have less traffic. In order to evaluate the performance of our model, it is necessary to compare it with other models: the second model we used is the artificial neural networks.

5.4.1.2 Model based on the ANN

We used artificial neural networks to predict the rating (score) of a trip. Let's consider that our features: trip_ID, user_ID, departure address, arrival address, latitude, longitude, departure date, arrival date, price, distance, duration, number of seats and means of transport are respectively X1, X2, X3,..., X13.

In the first implementation, (RandomForest based model), we encountered a gradient exploding problem. Our model was unable to complete processing, the problem produced when large error gradients accumulate and cause very large updates to the neural network model weights during training. When the magnitudes of the gradients accumulate, an unstable network is likely to occur, which can lead to poor prediction results. The explosion of gradients can lead to problems in the training of artificial neural networks and the learning cannot completed. The values of the weights can also become very large and lead to what we call Nan values. Nan values, which do not represent a number, but they represent undefined or unpresentable values which negatively influences the training of the model.

We cite three techniques we used in our project for solving the Exploding Gradients problem:

- Redesign the network model;
- Regularize the weights;
- Adding a Dropout layer.

After several adjustments, we achieved an accuracy of 93%. To be sure, of the obtained results, we displayed the confusion matrix (figure 8) which measures the quality of the model. Then each row corresponds to a real class and each column corresponds to a predicted class. In the first settings, we found that the model did not predict well the classes 0 and 1 with an accuracy equal to 20% for the class 0 and 50% for the class 1. However, after increasing the volume of the dataset we could get the desired results with an accuracy equal to 96% for class 0 and 90% for class.

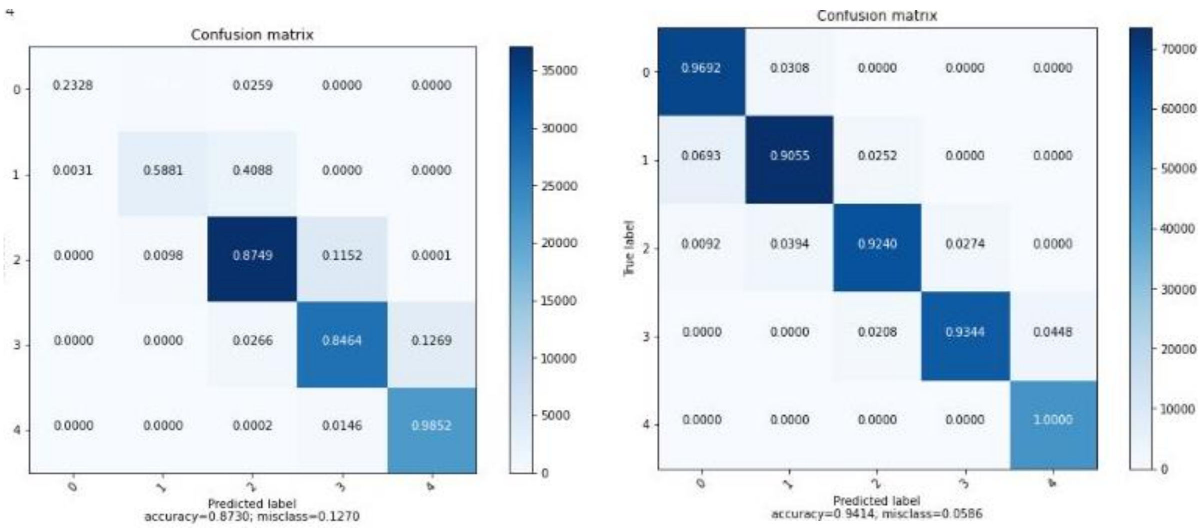

Figure 8: Visualization of confusion matrices before and after increasing the dataset.

In the content-based approach, we evaluated two models, the first Random Forest model allowed us to determine the impact of each feature on the algorithm, and the second Deep Learning ANN model allowed us to predict path scores.

5.4.2 Collaborative Filtering

The second approach we used is collaborative filtering in order to recommend a route to a user based on similar users.After the benchmark, we found that KNN Basic, which takes the score of each user, is the algorithm that gave the best value of Mean Absolute Error (MAE), which is a measure to calculate the difference between the actual scores and the predicted scores.

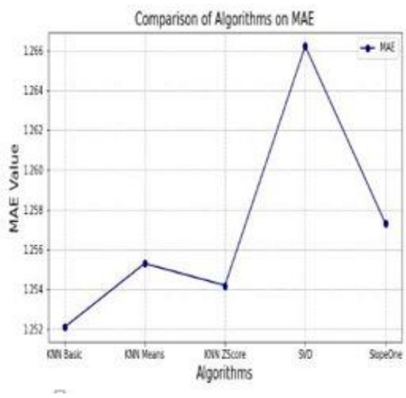

Figure 9: Visualization of benchmark results

5.5 Recommendation

After evaluating the two content-based approaches and collaborative filtering, we opted for the ANN algorithm of the content-based approach for recommending optimized routes with sorting of the predicted scores from best scored to worst scored. Therefore, we developed a test model function to perform the study process. We give our function the departure city, the arrival city and the non-preprocessed date and it returns all the routes meeting the query with a score sort.

	Ville de depart	Ville d'arrivée	Distance	prix	weekday_name_dep	Hour_dep	minute_dep	weekday_name_arr	Hour_arr	minute_arr	hour_duree	Moyen_de_transport	nombre_place	route	rating
0	RABAT	FES	275.181	68.7952	Wednesday	16	12	Wednesday	19	4	2	Voiture	1	autoroute	4
1	RABAT	FES	275.181	68.7952	Wednesday	12	54	Wednesday	15	46	2	Voiture	1	autoroute	4
2	RABAT	FES	275.181	77.0507	Wednesday	9	41	Wednesday	12	33	2	Voiture	1	autoroute	4
3	RABAT	FES	275.181	68.7952	Wednesday	9	6	Wednesday	11	58	2	Voiture	1	national	2
4	RABAT	FES	275.181	68.7952	Wednesday	13	4	Wednesday	15	56	2	Voiture	1	national	2
5	RABAT	FES	275.181	55.0362	Wednesday	18	24	Wednesday	21	16	2	Voiture	1	national	2
6	RABAT	FES	275.181	77.0507	Wednesday	7	29	Wednesday	10	21	2	Voiture	1	national	2
7	RABAT	FES	275.181	77.0507	Wednesday	13	33	Wednesday	16	25	2	Voiture	1	national	2
8	RABAT	FES	275.181	55.0362	Wednesday	22	45	Thursday	1	37	2	Voiture	1	national	1

Figure 10: Recommendation for best results

6 Conclusion

Summarizing what has already explained, the project aims to design and implement a recommendation system of the best monumental routes for the passengers. This recommendation system will allow passengers to choose the most optimal routes by avoiding traffic jams, timetables and unsafe routes. Thus, the user must fill in his city of departure, his city of arrival, the date and time of departure. Then, the system searches in the database the races corresponding to this trip, then, it applies the predictive model to obtain the results, which will sorted by order of importance.

The first part of the project on which we elaborated a monomodal recommendation, but the second part of the project will evoke the principle of multimodality that consists in recommending to a passenger more than one means of transport on the same journey. This approach requires all the implicit data of the users such as the number of visits, the number of clicks, and the use of the graph theory is to consider finding the optimal combination of the means of transport.

7 Références

- Abdelwahab, H.; Abdel Aty, M. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transp. Res. Rec. **2001**, 1746, 6–13.
- Adeniyi ;Yongquan .(2014) Automated Web Usage Data Mining and Recommendation System Using K-Nearest Neighbor (KNN) Classification Method | Elsevier Enhanced Reader ».

- Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2019). A review on deep learning for recommender systems : Challenges and remedies. Artificial Intelligence Review, 52(1), 1-37.
- Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley (1999).
- Brownlee, J. (2017, décembre 17). A Gentle Introduction to Exploding Gradients in Neural Networks. Site de Machine Learning Mastery. [En ligne]. Consulté 19 juin 2020, à l'adresse (https://machinelearningmastery.com/exploding-gradients-in-neural-networks/).
- Burke. 2002 Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction. 12, 4, 331-370.
- Dameri, R. P., & Rosenthal-Sabroux, C. (2014). Smart city and value creation. In Smart city (p. 1–12). Springer.
- Edge, S., Boluk, K., Groulx, M., & Quick, M. (2020). Exploring diverse lived experiences in the Smart City through Creative Analytic Practice. Cities, 96, 102478.
- Florea, A., & Berntzen, L. (2017). Green IT solutions for smart city's sustainability [5th Smart Cities Conference slides]. Retrieved from (http://administratiepublica.eu/smartcitiesconference/2017/files/SSC05-PP/ DAY%202/Green%20IT%20solutions%20for%20 smart%20city%20sustainability.pdf).
- Hayes, B. (2019). Programming Languages Most Used and Recommended by Data Scientists. Site BuisnessBroadway. [En ligne]. Consulté 26 juin 2020, à l'adresse (https://businessoverbroadway.com/2019/01/13/programming-languages-most-used-and-recommended-by-data-scientists/).
- Jayashree Nair. (2016).Random Forest Approach for Rating–based Recommender System AMRITA · Department of Computer Science· Master of Engineering.

- J.S. Breese, D.Heckerman, and C.Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Conference on Uncertainty in Artifical Intelligence, 1998.

- Karballaeezadeh, N., Zaremotekhases, F., Shamshirband, S., Mosavi, A., Nabipour, N., Csiba, P., & Várkonyi-Kóczy, A. R. (2020). Intelligent Road Inspection with Advanced Machine Learning; Hybrid Prediction Models for Smart Mobility and Transportation Maintenance Systems. Energies, 13(7), 1718.

- Lei, Vahid N., and Philip S. Y. Joint deep modeling of users and items using reviews for recommendation. WSDM, 2017.
- Lewald, A. (2017). Unlocking Mobility's Potential. How to Make Mobility Smarter and Cleaner. World News –Climate Change Te New Economy. Retrieved from (http://climatechange-theneweconomy.com/sustainable-transport-kapsch/).
- Nempanu, F., Schlingensiepen, J., buretea, D., & Iordache, V. (2016). Mobility as a service in smart cities. Responsible Entrepreneurship Vision, Development and Ethics, 425.
- Ranchordas, S. (2020). Smart Mobility, Transport Poverty, and the Right to Inclusive Mobility (SSRN Scholarly Paper ID 3520525). Social Science Research Network.
- Resnick, P. and Varian, H. R. 1997. Recommender systems. Communications of the ACM, 40(3), 56-58.
- Sachin Prabhu, ThandapanI (2019).Recommendation Systems : Collaborative Filtering using Matrix Factorization—Simplified. Site Medium. [En ligne]. Consulté 19 juin 2020, à l'adresse (https://medium.com/sfu-cspmp/recommendation-systems-collaborative-filteringusing-matrix-factorization-simplified-2118f4ef2cd3)
- Van Mead, N. (2017). Uber for bikes: how "dockless" cycles flooded China – and are heading overseas. Te Guardian. Retrieved from (https://www.theguardian.com/cities/2017/mar/22/bike-wars-docklesschina-millions-bicycles-hangzhou).
- Williams, J.; Li, Y. A case study using neural networks algorithms: Horse racing predictions in Jamaica. In Proceedings of the International Conference on Artificial Intelligence (ICAI 2008), Las Vegas, NV, USA, 14–17 July 2008.
- Xiangnan, Lizi L., Z. Hanwang, Xia H. Liqiang, N., and C. Tat-Seng. Neural collaborative filtering. WWW, pages 173—-182, 2017.
- Yongfeng, A. Qingyao, and C. Xu, C.and Bruce. Joint representation learning for top-n recommendation with heterogeneous information sources. CIKM, pages 1449—1458, 2017.
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019a). Deep Learning based Recommender System : A Survey and New Perspectives. ACM Computing Surveys, 52(1), 1-38.