

PAPER • OPEN ACCESS

Travel Time Estimation for Destination In Bali Using kNN-Regression Method with Tensorflow

To cite this article: Murni *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **854** 012061

View the [article online](#) for updates and enhancements.

You may also like

- [EPSPatNet86: eight-pointed star pattern learning network for detection ADHD disorder using EEG signals](#)
Dahiru Tanko, Prabal Datta Barua, Sengul Dogan et al.
- [Colorimetry-based and smartphone-assisted machine-learning model for quantification of urinary albumin](#)
Sunita Bhatt, Sunil Kumar, Mitesh Kumar Gupta et al.
- [A Comparison of Photometric Redshift Techniques for Large Radio Surveys](#)
Ray P. Norris, M. Salvato, G. Longo et al.

UNITED THROUGH SCIENCE & TECHNOLOGY



The Electrochemical Society
Advancing solid state & electrochemical science & technology

**248th
ECS Meeting**
Chicago, IL
October 12-16, 2025
Hilton Chicago



**Science +
Technology +
YOU!**

**Register by
September 22
to save \$\$**

REGISTER NOW

Travel Time Estimation for Destination In Bali Using kNN-Regression Method with Tensorflow

Murni, R Kosasih, A Fahrurrozi*, T Handhika, I Sari, D P Lestari

Computational Mathematics Study Center, Gunadarma University, 100 Margonda Raya Street, Pondok Cina, Depok, West Java, 16424, Indonesia

*E-mail: achmad_fahrurrozi@staff.gunadarma.ac.id

Abstract. On a tour activity, travel time estimation is needed so that the travel itinerary goes according to the plan. Travel time estimation is very important so we can estimate the time needed to arrive at the destinations in the travel itinerary. Therefore we need a method that can estimate travel time from one place to another. In this study, we propose the k-Nearest Neighbors Regression (kNN-Regression) method with Tensorflow to construct an estimation model. The proposed number of features in our estimation model is 8 features, i.e. zone information, time information, day information, weather information, temperature information, wind speed information, humidity information, and precipitation information. The data obtained from travel information from Ngurah Rai airport to Kuta Beach using GPS and weather information using weather application in real-time. We divide our data into two groups: a historical group consisting of 177 data and a testing group consisting of 51 data. In the testing stage, kNN-Regression will find the historical data closest to the testing data, so that the estimation value of the travel time of some testing data is not much different from the value of the nearest historical data. As a result, our proposed model gives the Mean Absolute Error (MAE) of 2.196078, Root Mean Square Error (RMSE) of 2.977036294 and accuracy rate 88.1819%.

Keyword: KNN-Regression, Features, MAE, RMSE, Accuracy rate

1. Introduction

Travel time estimation is something that is suggested to be done before going on a trip. One of the problems in travel time estimation is traffic jams. If a traffic jam occurs it will cause a longer travel time. So that we need a method that can estimate travel time from one place to another place. Travel time estimation is very important so that we can estimate the time needed to arrive at the destination. This estimation can also be used by travel agents such as travel agents to be able to estimate when the departure time and which route will be chosen. So that the travel agent can estimate how many destinations you want to visit in one day.

In planning a trip, there are several features that can affect travel time. For example, are time information. Departure time will affect travel time, for example, is the number of people who go to work in the morning by using a vehicle will cause traffic jams, or the number of people who want to go on vacation on a holiday will also cause traffic jams. The other features are weather conditions such as rain can cause the vehicle rate to be slower because the road becomes more slippery so that it can cause a traffic jam. The next feature is the temperature, if the temperature in the area is very extreme, it will cause people to do not want to leave the house so that it can cause traffic to be smooth.



Several studies have been carried out such as D'Angelo et al. Predicting travel time using a nonlinear time series model and arguing that models that only use speed data are superior to multivariable models that use speed, occupancy, and volume of data obtained from loop detectors [1]. Next Kwon et al predicted travel time on the freeway by using flow and occupancy data from a single loop detector and historical travel time information [2]. They use linear regression and tree-based methods to show that the current traffic data collected from the loop detector is a good predictor for the near future (up to 20 minutes). Zheng Wang uses a regression problem to predict vehicle travel time based on floating car data by considering account factors such as spatial information, temporal information, traffic information and personal information [3]. However, this study only considers 4 features. Therefore in this research, we will find other features that can affect travel time. We proposed a model with 8 features such as zone information, time information, day information, weather information, temperature information, wind speed information, humidity information, and precipitation (rainfall) information by using KNN regression method with TensorFlow.

2. Literature Review

This section describes the theory of travel time estimation using KNN regression with Tensorflow.

2.1. K Nearest Neighbor (KNN)

K Nearest Neighbor (KNN) algorithm is a method of classifying objects based on their similarity to other objects. The KNN is the classification technique without having to know about the distribution of the data [4-7]. In this method, each sample should be classified similarly to the surrounding samples. Therefore, if the classification of a sample is unknown, then it could be estimated by considering the classification of its nearest neighbor samples [7]. The Nearest Neighbor rule (NN) is the simplest form of KNN when $K = 1$ [7].

For example, suppose there is one object that wants to be classified using training data. The first stage is to calculate all distances from the object to all training data. The distance with the smallest value corresponds to the sample in the training set closest to the unknown data. Therefore, the unknown sample may be classified based on the classification of this nearest neighbor [6-10].

The stages of KNN algorithm can be seen in Figure 1.

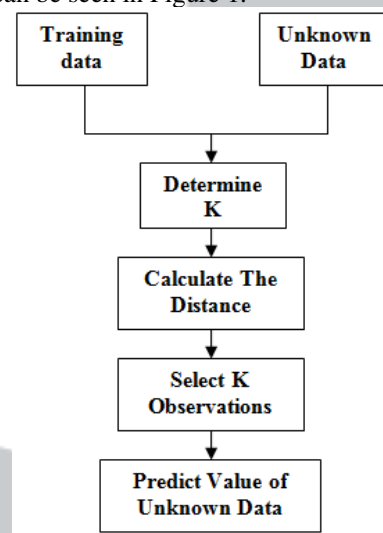


Figure 1. KNN Algorithm.

In Figure 1, we determine the value of K and calculate the distance from the unknown object to all training data. After that, select K observations in the training data closest to the unknown data. The last

step is to predict the value of unknown data by using the value of the K-nearest neighbor object that was selected in the previous step.

2.2. Tensorflow

TensorFlow is a machine learning system that operates at large scale and in heterogeneous environments. TensorFlow is a library from a language of python that used for numerical computation that makes machine learning faster and easier. TensorFlow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array or *tensor*.

TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations. Best of all, TensorFlow supports production prediction at scale, with the same models used for training.

3. Research Methodology

This section describes the stages of the method of research that has been done. The stages of research can be seen in Figure 2.

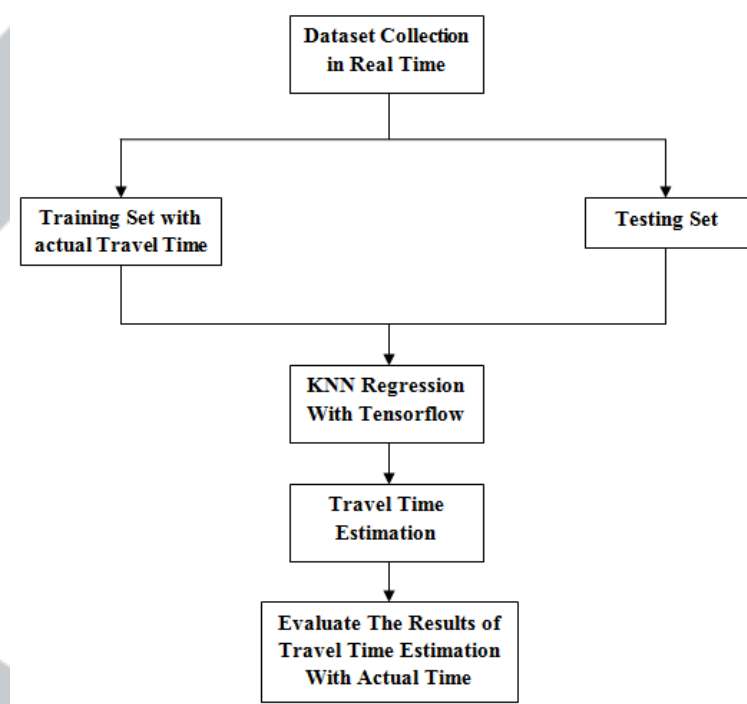


Figure 2. General frameworks for this research.

In Figure 2, the first step in estimating travel time is the collection of datasets in real-time. The data is divided into 2 namely training data (historical data) with actual travel time and testing data used to estimate travel time. To estimate travel time we use the KNN Regression method with tensorflow. Next, we evaluate the estimation results by using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Level of Accuracy (Acc).

3.1. Dataset Collection

In this study, travel information is taken through GPS and the weather information dataset is taken through weather application in real-time. The sample of travel information from GPS and weather information can be seen in Figure 3 and Figure 4.

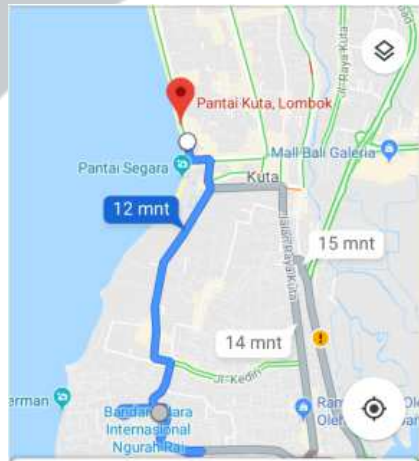


Figure 3. Travel information from GPS in real-time



Figure 4. Weather information in real-time

The dataset is divided into two namely 177 training data (historical data) and 51 testing data. The training data was collected manually through the application from 25 July 2019 until 1 August 2019 in the Bali region especially from Ngurah Rai Airport to Kuta beach tourist attraction. After that testing data was collected manually through the application from 9 August 2019 until 12 August 2019. Data collected is

- **Zone Information**
Zone information is one of the features that can be used to estimate travel time. In this study, there are 3 zones that can be passed from the Airport to Kuta Beach. For example, different zones will produce different travel times.
- **Time of The Day Information**
Time information can affect travel time estimation. That is because at certain times there is a pattern of traffic. For example in the morning, many people who do activities so that traffic jams occur
- **Day of The Week Information**
Day information can affect travel times estimation. For example, on weekends many people go on vacation to attractions, causing traffic jams
- **Weather Conditions**
Weather conditions can also affect travel time estimation. For example, if extreme weather occurs, such as extreme heat, it will cause people to stay at home so there is no traffic jam
- **Temperature Information**
Temperature conditions can affect travel time estimation. Such extreme temperatures tend to keep people inside the house so that traffic congestion decreases and causes travel times to be faster than usual
- **Wind Speed Information**
Wind speed can affect vehicle speed. If the wind speed is very high, it will cause the vehicle speed to decrease, causing traffic jams
- **Humidity Information**
The higher the humidity the air will cause the level of pollution is higher so as to make people stay in the house so that traffic density decreases and causes travel times to be faster than usual

- Precipitation Information
if the precipitation (rainfall) is very high it will cause people to stay inside the house so that traffic density decreases and causes travel times to be faster than usual.

3.2. KNN Regression With Tensorflow

In addition to classification, K Nearest Neighbors (KNN) can also be applied for regression. The stages of KNN regression can be seen in Figure 5.

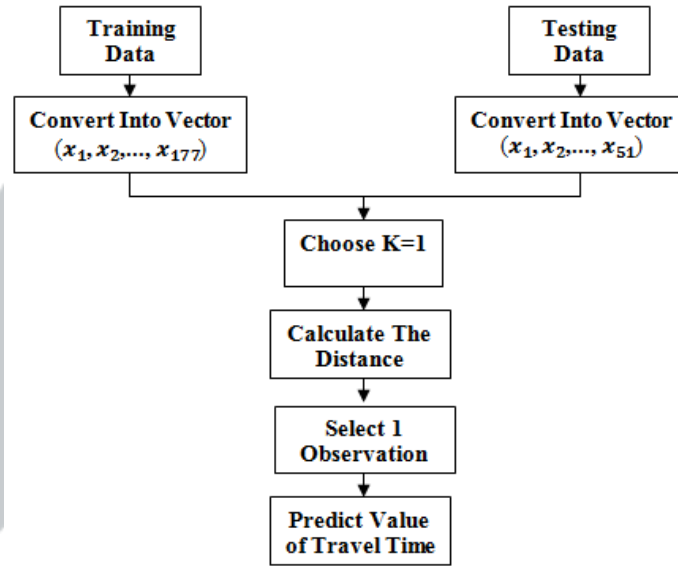


Figure 5. KNN Regression Algorithm.

In Figure 5, we use 8 features to predict travel time from Ngurah Rai Airport to Kuta Beach. Furthermore, each data that has 8 features is assumed to be a vector x_i with dimension 8. In this study, we use 177 training data (Historical Data). Therefore there are 177 vectors $(x_1, x_2, \dots, x_{177})$. Each vector represents a point in 8 dimension space.

The next stage is to predict travel time by using 51 testing data. After that, we use the KNN Regression to predict travel time. In this study, the K used was 1, in other words, 1 was chosen closest to the training data. Furthermore, calculating the distance from every testing data to all training data using Manhattan distance. The Formula of Manhattan Distance can be seen in equation (1).

$$MD = \sum_{i=1}^k |a_i - b_i| \quad (1)$$

Where a_i is testing data and b_i is training data.

KNN will find the training data (historical data) closest to the testing data so that the predicted travel time is not much different from the value in the closest historical data.

3.3. Evaluate

After estimating the travel time by using the KNN regression, the next step, we evaluate the results of travel time estimation using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Level of Accuracy (Acc).

The formula of Root Mean Squared Error (RMSE) [7] can be seen in equation (2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (2)$$

and the formula of Mean Absolute Error (MAE) [11] can be seen in equation (3).

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (3)$$

and the formula of Accuracy (Acc)[11] can be seen in equation (4).

$$Accuracy = \left(1 - \frac{\sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{N} \right) \times 100\% \quad (4)$$

Where y_i is actual travel time, \hat{y}_i is the result of travel time estimation and N is the number of dataset testing

4. Experimental Result and Discussion

In this study, we use two models to estimate travel time. The first model is a model with 5 features (model 1) and the second model is a proposed model with 8 features.

Model 1 can be seen in equation (5)

$$y = \sum_{i=1}^5 x_i + \varepsilon \quad (5)$$

Where x_1 = Zone information, x_2 = Time of The Day, x_3 = Day of The Week, x_4 = Weather Conditions, and x_5 = Temperature Information

Futhermore, proposed model can be seen in equation (6)

$$y = \sum_{i=1}^8 x_i + \varepsilon \quad (6)$$

Where x_1 = Zone information, x_2 = Time of The Day, x_3 = Day of The Week, x_4 = Weather Conditions, x_5 = Temperature Information, x_6 = wind speed information, x_7 = Humidity Information, and x_8 = precipitation information

The next stage is to find the predicted value of two models by using the KNN regression method. After that, we evaluate two models by calculating the value of MAE, RMSE, and Accuracy which can be seen in Table 1.

Table 1. Evaluate Model

| | MAE | RMSE | Accuracy |
|----------------|----------|-------------|------------|
| Model 1 | 2,254902 | 3,0901488 | 86,2103 % |
| Proposed Model | 2.196078 | 2,977036294 | 88,1819 %. |

In Table 1, there are two models being compared, namely a Model 1 with 5 features and a Proposed Model with 8 features. From table 1 it can be seen that the value of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) of the proposed model is smaller than the value obtained using a model with 5 features and for the level of accuracy of the proposed model has a higher value compared to the accuracy of model 1 with 5 features. This shows that the Proposed Model is better than Model 1 with 5 features.

5. Conclusion

Travel time estimation is needed by someone or a travel agent who wants to travel. Therefore we need a method that can estimate travel time from one place to another. In this study to estimate travel time, we propose using the K Nearest Neighbors (KNN) regression method by considering features such as zone information, time information, day information, weather information, temperature information, wind speed information, Humidity information, and precipitation information. Data retrieval of these features is obtained by using GPS data and weather information in real-time through the weather application. The data is divided into two namely 177 training data (historical data) and 51 testing data. In testing, KNN will find the historical data closest to the testing data so that the predicted travel time is not much different from the value in the closest historical data. In this study, we use two models namely a model with 5 features and a proposed model with 8 features. From the evaluation results using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and level of Accuracy, it can be seen that the proposed model is better than the model 1 with 5 features. For the next experimental research, we will add a sample of training to get higher accuracy.

Acknowledgment

This research was supported by National Competitive Research Grant 2019 in Penelitian Terapan (PT) scheme, DRPM KEMENRISTEKDIKTI Indonesia to the Gunadarma University.

Reference

- [1] D'Angelo M P, Al-Deek H M, and Wang M C 1999 Travel-Time Prediction for Freeway Corridors. *J. Transp. Res. Board* **1676** 184–191.
- [2] Kwon J Coifman B, and Bickel P 2000 Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop-Detector Data *J. Transp. Res. Board* **1717** 120–9.
- [3] Wang Z, Fu K and Ye J 2018 Learning to Estimate the Travel Time *Proc. Int. Conf. on Knowledge Discovery & Data Mining (London)* vol 24 (New York : USA ACM) p 858-66.
- [4] Devroye L 1981 On the equality of Cover and Hart in nearest neighbor discrimination *IEEE Trans. Pattern Anal. Mach. Intell.* **3** 75-8.
- [5] Devroye L, Györfi L, Krzyżak A and Lugosi G 1994 On the strong universal consistency of nearest neighbor regression function estimates *Ann. Statist* **22** 1371–85.
- [6] Devroye L and Wagner T J 1982 Nearest neighbor methods in discrimination, In Classification Pattern Recognition and Reduction of Dimensionality *Handbook of Statistics* vol 2 (Amsterdam : Elsevier) p 193–7.
- [7] Imandoust S B And Bolandraftar M 2013 Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background *J. Eng. Res. and App.* 3 (5), 605-10
- [8] Domeniconi C, Peng J and Gunopulos D 2002 Locally adaptive metric nearest-neighbor classification *IEEE Trans. Pattern Anal. Mach. Intell.* 24(9) 1281–5.
- [9] Dudani S A 1976 The distance-weighted k-nearest neighbor rule *IEEE Trans. Sys. Man. Cyber.* **6** 325-7.
- [10] Enas G G and Choi S C 1986 Choice the smoothing parameter and efficiency of K-Nearest Neighbor classification *Comp. Maths. Apps.* 12(2) 235-44.
- [11] Kumar B A, Jairam R, Arkatkar S S and Vanajakshi L 2019 Real time bus travel time prediction using k-NN classifier *Int. J. Transp. Res.* 11(7) 362-72.