

Data Mining in the bike sharing dataset

by Mingyang Zhu

Submission date: 09-Mar-2023 05:37AM (UTC+0000)

Submission ID: 198837680

File name: 40847_Mingyang_Zhu_Data_Mining_in_the_bike_sharing_dataset_482892_324183279.docx (19.24K)

Word count: 566

Character count: 3095

Data Mining in the bike sharing dataset

– Summary Document

The origins of the data

Washington State has passed the Capital Bikeshare Data License Agreement, and the government has installed cycling recorders on shared bicycles throughout the city. According to the data from the recorders, researchers can get start date and time, end date and time, start point, end point, the bike ID number and the user's membership type of each ride. On the basis of the original data, Dr. Roberto Frias of the University of Porto, laboratory of Artificial Intelligence and Decision Support (LIAAD), got the weather and the holidays information from <http://www.freemeteo.com> and <http://dchr.dc.gov/page/holiday-schedule>, he classified them into season, holiday, weather, temperature, humidity, wind speed and other factors, and merged them into the original dataset[1]. Finally, bike sharing dataset is obtained.

The Dataset

The bike-sharing rental system has developed quite mature now. It has fully realized automation from rental to return. Users can easily rent bikes by themselves and return them at a return point near their destination. As mentioned in the Capital Bikeshare Data License Agreement, the original intention of the project is to provide information about where people ride, when they ride, and how far they ride to engineers, scholars and anyone else who may be interested in it for analysis and research. The bike sharing dataset contains all types of bike-sharing usage information in the Washington area during 2011 and 2012, plus other attributes such as season, month, temperature, humidity, wind speed, whether it is a working day, whether it is a holiday, etc., which information was integrated by Dr. Roberto Frias. In general, there are 17,389 instances in total and 16 attributes per instance.

Suitability for analysis

The bike sharing dataset has 17389 examples, such a large number of examples makes the conclusions analyzed more applicable. In addition, on the basis of the original riding time and date, Dr. Roberto Frias added elements about weather and holiday schedule. These later added factors make the analysis more targeted and diversity, so that analysts have the opportunity to study the association between these factors and riding time, and can further compare the influence weight of different factors and under what circumstances people are more inclined to rent shared bicycles.

Potential research questions

1. In what month do people tend to rent bicycles?
2. In what season do people tend to rent bicycles?
3. In what weather do people tend to rent bikes?
4. At what temperature do people tend to rent bikes?
5. In what humidity do people tend to rent bikes?
6. In what wind speed do people tend to rent bicycles?
7. Does the rental demand for shared bikes increase or decrease in 2012 compared with 2011?
8. Who has the greatest influence weight among different factors?

Researcher background

1. As a practical user of shared bikes, the author has personally experienced the influence of

various factors on users' rental time, which is helpful for the author to analyze the research conclusions.

2. The author is actively learning related knowledge of data mining, including but not limited to Tableau and Jamovi software.

References

- [1] UCI Machine Learning Repository, "Bike Sharing Dataset", Online source: [UCI Machine Learning Repository: Bike Sharing Dataset Data Set](#) (accessed 9/3/2023)
- [2] Fanaee-T, Hadi, and Gama, Joao, 'Event labeling combining ensemble detectors and background knowledge', Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg,

Video link

<https://youtu.be/carcWDyweyc>

Data Mining in the bike sharing dataset

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

Instructor

66/100

PAGE 1

PAGE 2

GRADING FORM: DATA MINING/ VISUALISATION

MINGYANG ZHU

66

KNOWLEDGE AND UNDERSTANDING

Dataset Suitability (What features of the dataset lend themselves to data mining? Has it been mined previously? Why does the student feel that they have the contextual background to work with this data?)



Dataset description and hosted location are listed in submission. Sample size is sufficient for future analysis.

Type of data within appraised dataset is not explained sufficiently however.

19

COGNITIVE AND INTELLECTUAL

Objectives (What does the student plan to investigate/expect to find? Are these SMART?)



Potential questions are highlighted but student own expectations of future analysis of collated dataset is not made explicit.

6

PRACTICAL APPLICATION SKILLS

Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved?)



Very limited explanations provide about chosen dataset, its viability for future analysis; data types ; and sample usefulness.

35

TRANSFERABLE SKILLS FOR LIFE

Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance)



Suitably formatted submission with clear headings and supporting information used for support an understanding of assignment requirements.

6