RHEIN-WAAL UNIVERSITY OF APPLIED SCIENCES

Machine Learning Based Restaurant Location Analysis in Hamburg

by

Hung Viet Hu, Sachin Kumar, Vincent Meyer zu Wickern

in the

Faculty of Communication and Environment

Geoinformatics, WS 2018/19

February 2019

I, Vincent Finn Alexander Meyer zu Wickern, declare that this thesis titled, 'Predictive Maintenance' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this paper has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this paper is entirely my own work.

- I have acknowledged all main sources of help.

- Where the paper is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

RHEIN-WAAL UNIVERSITY OF APPLIED SCIENCES

Abstract

Faculty of Communication and Environment

Geoinformatics, WS 2018/19

by Hung Viet Hu, Sachin Kumar, Vincent Meyer zu Wickern

Contents

Abbreviations

MetaVer© MetadatenVerbund

Chapter 1

Introduction

Chapter 2

Elements of a restaurant location analysis

Chapter 3

Data Sources

3.1 Transparency Portal Hamburg

As the first German federal state, Hamburg enacted a transparency law on October 6, 2012 [1]. Opposed to a right to request information, which all citizens had until this date, a new duty to inform the public was laid upon the state's administration offices. All information that would fall under this law, now had to to be published in a freely available standard format on a centered storage of information. The single pieces of information, which would fall under the law, varied highly in precision and the comprehensive term of "geodata" was requested opposed to precise datasets of geodata. A legal interpretation was worked out for all requested points and a plan for the release of geodata was designed consisting of the basic data for measurement admistration and the technical geodata for special administration offices. The transparency law granted a period of two years for the technical implementation.

In October 2014, the "Transparency Portal" (http://transparenz.hamburg.de/) as the major component of the implementation of the transparency law was released [1]. With this portal, the Hamburg citizens have a multitude of data and documents available that was prior only available to Hamburg's administration. One important focus was the release of geodata that was even before the law in preparation for an "Open GeoData" model. In this "Open GeoData" model, geodata was split into two groups of data sets, one group extractable with little effort, but free to the public and expected with a high use, and another group with expected high demand and high revenue on the sale of this data. For this second group of datasets, more effort with new measurements had to be arranged. With the transparency law in place, all datasets were merged into the Transparency Portal and yielded a much higher download count than the count of dataset sales before the portal was active. The Transparency Portal uses a standardized meta data repository called the MetadatenVerbund (MetaVer©) in collaboration with other German federal states.

3.2 Yelp

Chapter 4

Analysis Methods

4.1   Random Forest Regression

4.2   Performance Measures

Chapter 5

Data Extraction

5.1   Yelp Restaurant Data Extraction

5.2   Hamburg District Map

Since usually cities raise important figures in aggregation per admistrative area, in Hamburg being the single city districts, these administrative areas should be imported into QGIS to be able to link single restaurants to an administrative area and hence figures that could be important as dependent variables to predict restaurant success. These administrative areas are taken from a dataset of the Transparency Portal called "ALKIS Verwaltungsgrenzen Hamburg" [2].

5.3   Social Values

5.4   Proximity to Water

Chapter 6

Machine Learning

6.1 Exploratory Data Analysis

6.2 Data Preprocessing

6.2.1 Handling of Missing Values

6.2.2 Feature Subset Selection

6.2.3 Dmensionality Reduction

6.3 Data Analysis

Chapter 7

Results and Discussion

7.1 Results

7.2 Discussion

Chapter 8

Conclusion

Bibliography

[1] Roswitha Murjahn and Sascha Tegtmeyer. Open data/transparenzportal hamburg–grundlagen, umsetzung, erfahrungen, auswirkungen. *Zfv–Z Für Geodäsie Geoinformation Landmanagement*, 5 (2016):330–335, 2016.

[2] Alkis verwaltungsgrenzen hamburg, 28.02.2018. URL http://suche.transparenz.hamburg.de/ dataset/alkis-verwaltungsgrenzen-hamburg8?forceWeb=true.