

## Using Twitter to assess the influence of urban environment on well-being in Lisbon population

Iuria Betco<sup>1</sup>, Luis Encalada-Abarca<sup>1,2</sup>, Ana I. Ribeiro<sup>3,4,5</sup>, David S. Vale<sup>6</sup>, Jorge Rocha<sup>1</sup>

<sup>1</sup>Institute of Geography and Spatial Planning, Associate Laboratory Terra, University of Lisbon, Portugal  
<sup>2</sup>Universidad Espíritu Santo, Ecuador  
<sup>3</sup>EPIUnit – Instituto de Saúde Pública, Universidade do Porto, Portugal  
<sup>4</sup>Departamento de Ciências da Saúde Pública e Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto, Portugal  
<sup>5</sup>Laboratório para a Investigação Integrativa e Translacional em Saúde Populacional (ITR), Portugal  
<sup>6</sup>Lisbon School of Architecture, University of Lisbon, Portugal



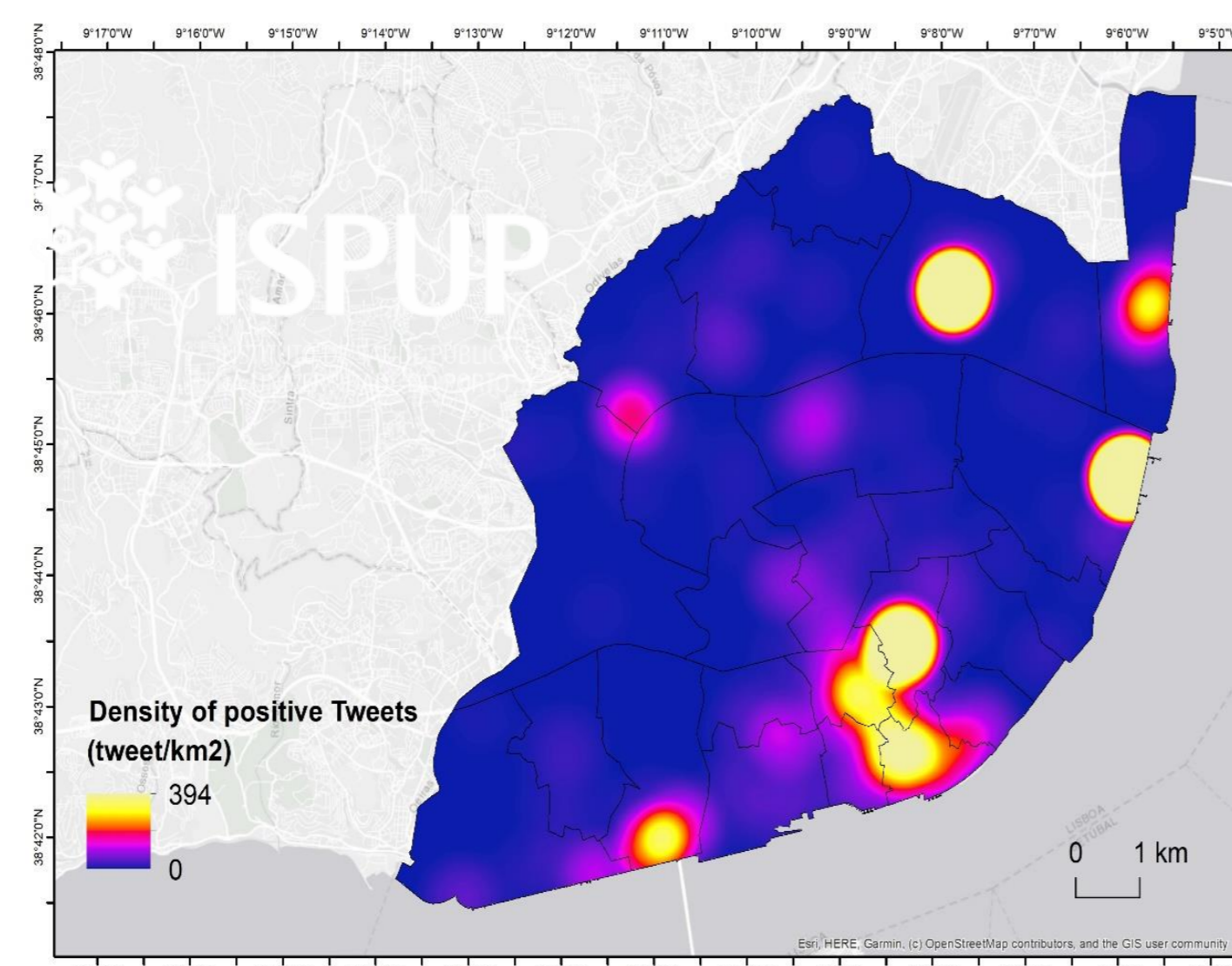
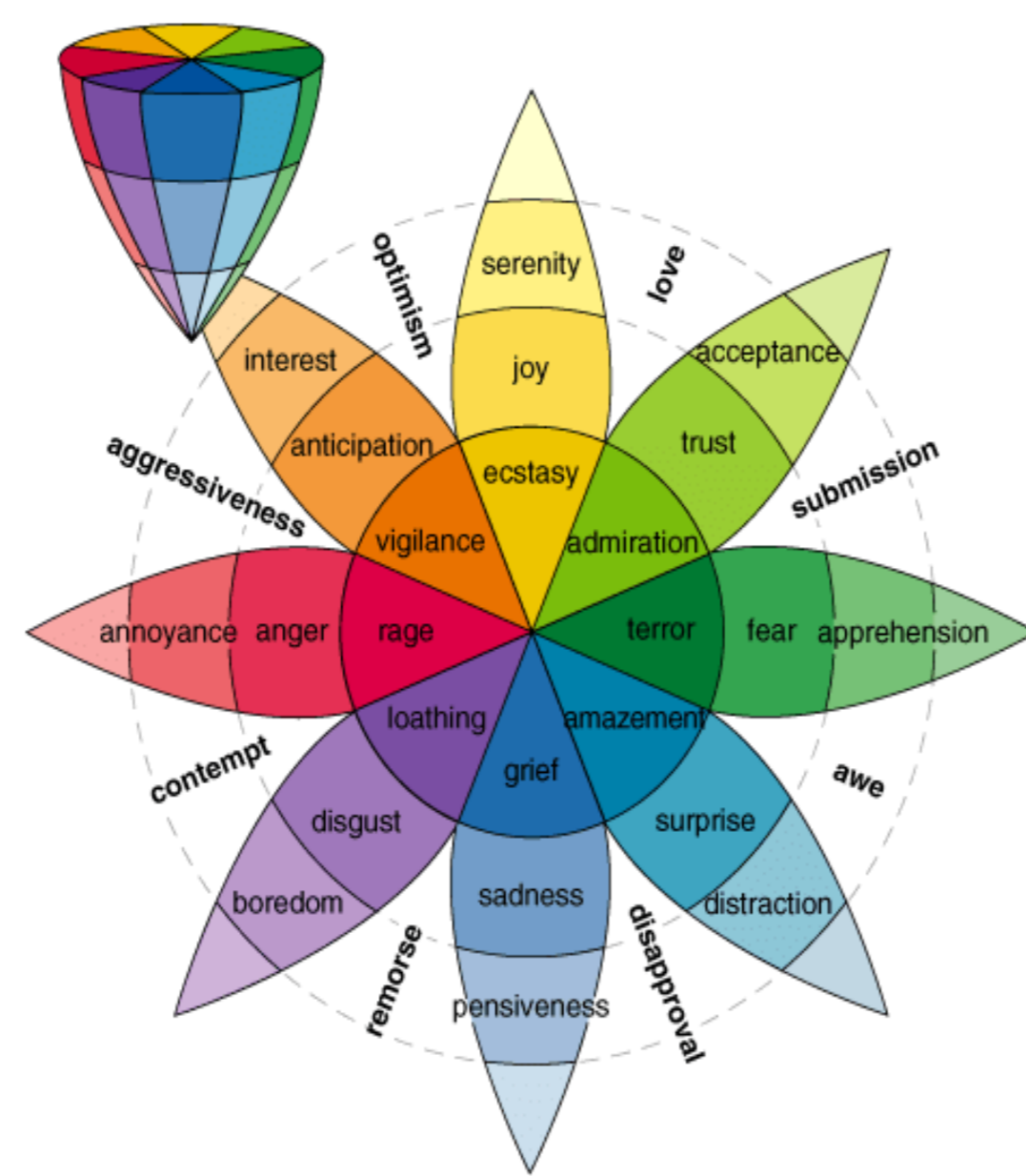
Jorge.rocha@campus.ul.pt

### INTRODUCTION

The shape and morphology of the city can be determinants of well-being, and it is essential to determine which factors can contribute to achieving this objective, in order to design the city in this sense. The transformation of urban space, simultaneously with the digitization of society, enables knowledge about how much individual well-being is influenced by the surrounding built environment. The content generated by internet users on social networks provides information about users and their environment, allowing access to their preferences, opinions and interactions. This study aimed to identify which features of the urban environment of the city of Lisbon affected well-being in 2019 through the use of georeferenced Twitter comments.

### BACKGROUND

The concept of well-being has a multifaceted nature that can be defined and measured in several ways. Well-being has two dimensions: i) Objective, an approximation measure based on basic human rights and needs; and ii) Subjective, measured by directly asking people how they think and feel about their own well-being. Inside the last one we have hedonic well-being as the most primitive and short-term, resulting from pleasure and pain prevention. A person with high subjective well-being has high levels of satisfaction, high positive sentiments and low negative sentiments.



### DATA

Twitter data allows for a more economical and dynamic approach compared to the usual survey and, more importantly, the ability to add an objective geometric dimension to the subjective analysis. In addition to spatial information, the content of tweets also provides information about users' opinions and feelings, allowing for a sentiment analysis.

The shape and morphology of the city can be determinants of well-being, and it is essential to determine which factors can contribute to achieving this objective, in order to design the city in this sense. The transformation of urban space, simultaneously with the digitization of society, enables knowledge about how much individual well-being is influenced by the surrounding built environment. The content generated by internet users on social networks provides information about users and their environment, allowing access to their preferences, opinions and interactions. This study aimed to identify which features of the urban environment of the city of Lisbon affected well-being in 2019 through the use of georeferenced Twitter comments.

### INDEPENDENT VARIABLES

**BUILDINGS AGE**

Determination of the weighted average age of the buildings  
**INE**

**URBAN MORPHOLOGY**

Relative Richness  
Diversity  
Dominance  
Fragmentation  
**COS**

**NOISE**

Global Noise Map (Lden) of the city of Lisbon  
**CML**

**POINTS OF INTEREST**

Determination of distance to equipment with influence on well-being  
**CML**

**GREEN AND BLUE PRINT**

Built-Up Index  
Non-Homogeneous Feature  
Vegetation Index  
Water Index  
**WorldView-2**

**POPULARITY**

Analysis of popularity of spaces in the city  
**Flickr**

**STREET ATTRACTIVENESS**

Determination of global solar radiation and street visibility  
**Lidar**

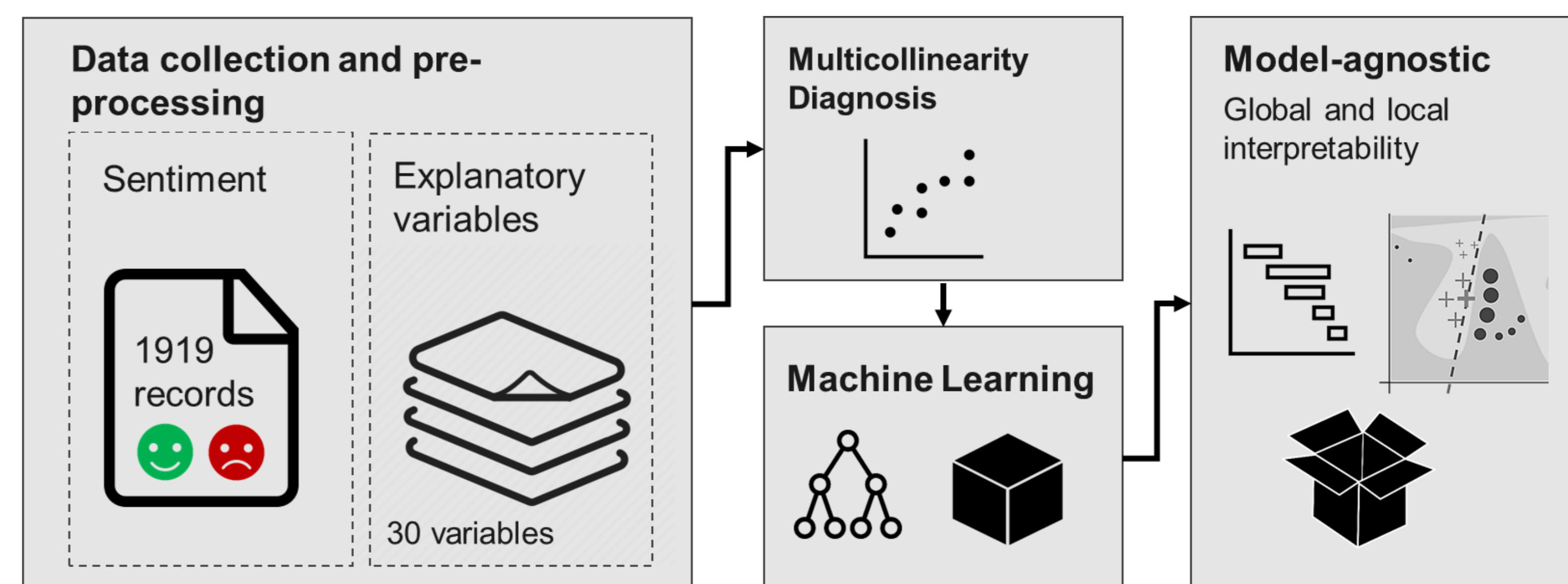
**POLLUTION**

Calculation of the daily average (9-19) of pollutant emission: PM<sub>10</sub>, NO<sub>2</sub> e O<sub>3</sub>  
**APA**

**CRIME**

Identify the most unsafe areas based on crime density  
**PSP**

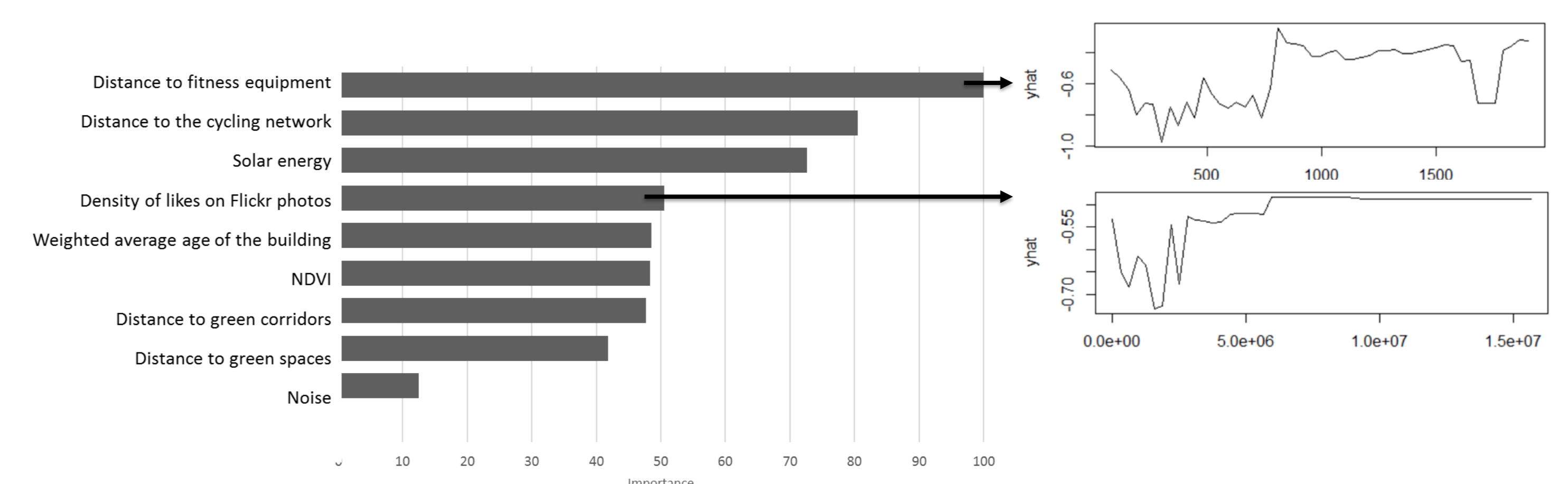
### METHODOLOGY



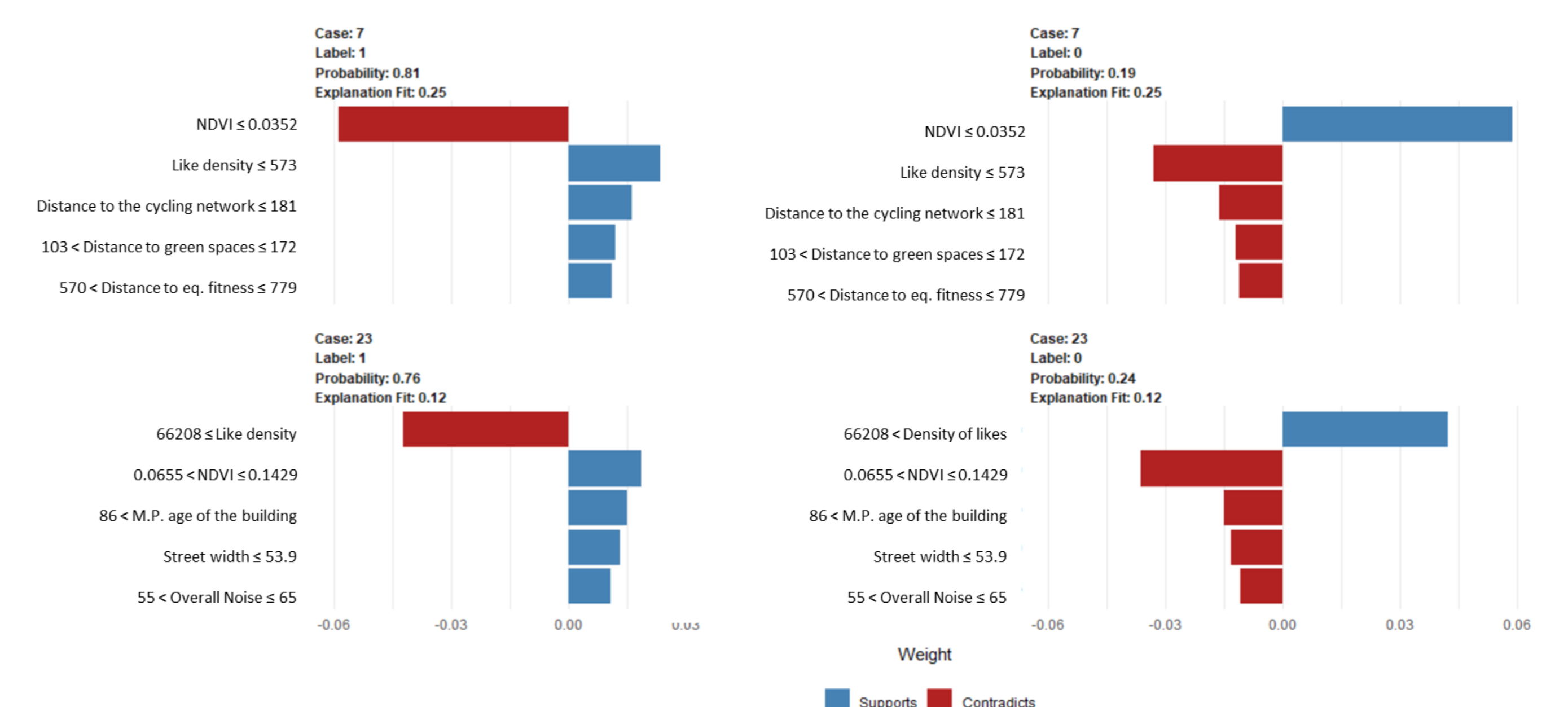
Algorithm	Accuracy	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value
rf	0.8125	0.8095	0.814	0.68	0.8974
glm	0.7188	0.2222	0.96124	0.73684	0.71676
nnet	0.7344	0.2222	0.9845	0.875	0.72159
knn	0.8021	0.8095	0.7984	0.6623	0.8957
xgboost	0.8021	0.8254	0.7907	0.6582	0.9027

### RESULTS

Permutation Feature Importance (PFI)  
Partial Dependence Plot (PDP)



### Local Interpretable Model-agnostic Explanations (LIME)



### CONCLUSIONS

Twitter data is a good proxy for sentiment analysis. The Portuguese are happy in spaces associated with leisure and consumption: "People does like to shop". Agnostic models and especially those based on game theory play a key role in this type of analysis. Emolex lexicon has some known limitations (punctuation, polarity, language). Using Syllabus can still be useful, despite the existence of other methods – helps to train the machine learning method. It may be useful to test other sentiment ranking techniques, such as Ontology-based and Machine Learning.