Digital discrimination detection in ridesharing services

Raquel T. de Paiva, Wendy S. Cataldo, Ana Cristina B. Garcia, Carlos E. De Mello

[1]Instituto de Informática – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
CEP 22.290-255 – Rio de Janeiro – RJ – Brazil

raquel.paiva@edu.unirio.br, wendy.cataldo@edu.unirio.br,

cristina.bicharra@uniriotec.br, mello@uniriotec.br

Abstract. *The introduction of technology has significantly transformed the urban transport industry, revealing social issues such as bias-driven trip cancellations. After a bibliographical review on how the topic is treated, we created an ontology and established the objective of analyzing digital discrimination, approaching it through the analysis of collective data, which can direct mechanisms to discourage discrimination in digital services. This study seeks to answer: RQ1: Is there evidence of digital discrimination in the shared transport service in the city of Rio de Janeiro? RQ2: Is it possible to identify the factors that lead to discrimination? RQ3: What are the key concepts related to detecting Digital Discrimination in a shared transport service?*

1. Introduction

Currently, urban mobility has become a major challenge in large centers. The increase in the number of cars and public transport that has been suffering with quality and quantity, is a persistent and multifaceted problem, that directly influences urban optimization with direct consequences on traffic flow and congestion [Batty 2012], and impact on the environment, in addition to having a direct relationship with the restriction of offers of legalized urban mobility services. and guaranteed by public authorities and an increase in irregular and non-legalized transport services.

In addition to these, the population suffers from low-quality public transport and individual transport services, such as taxis, which had high costs and also low quality of service, as they had a monopoly on this service. Given this scenario, in 2014 the ridesharing service arrived in Brazil, where the urban transport sector has been significantly transformed by technology, with transport applications playing a key role in simplifying and efficient passenger[Miroslav Tushev and Mahmoud 2020], causing new platforms to emerge, increasing competition, and allowing the population to have more transport options at affordable costs.

However, the ridesharing apps only enable the provision of a service in a peer-to-peer modality, that is, between a passenger and a driver, who in turn have biases, a phenomenon that undermines equality and accessibility of transport services [Ge 2018].

These biases can lead to a social problem which is discrimination. This brings us to a question: would ridesharing app available in Rio de Janeiro city, be immune to discrimination related to gender, race, age, able, class and religious, among other characteristics [Jorge Mejia 2020]? Does the ridesharing service available have other issues that we are not aware of?

The database used was Google Scholar and the search strings were:

- “discrimination prejudice bias ridesharing applications science computing”
- ”uber ”gender discrimination” source:Information source:Systems”
- ”uber ”gender discrimination” source:IEEE”

Filtering for the last 5 years, the search resulted in 1030 articles in the first query, 9 articles in the second, and 6 articles in the last.

A pre-selection was then carried out using the keywords found in the article's title and their origin and number of citations. After this pre-selection, the content of the abstract and conclusion were analyzed. If the selected article had a scope that fits the researched problem, which resulted in the full read analysis of 14 articles.

With this analysis, our main motivation is to answer the following research questions:

- **RQ1**: Is there evidence of digital discrimination in the ridesharing application used in Rio de Janeiro city?
- **RQ2**: It is possible to identify the factors that lead to discrimination?
- **RQ3**: What are the key concepts regarding digital discrimination detection in a ridesharing service?

The rest of this paper is organized as follows. Section 2 provides a background on discrimination. In section 3, we describe and present our ontology, followed by our methodology in section 4. Next in section 5, we present our study results and discussion. Section 6 addresses the conclusion and limitations of this study.

2. Background

In this section, we will present the main concepts for understanding the research problem and the techniques and approaches for analyzing the problem in this work.

To analyze the main complaints from users from one platform of ridesharing service and whether there is any factor of discrimination related to users of the service, it is first necessary to understand what discrimination is, how it occurs, and whether it can be reflected in digital services. By understanding how and when discrimination can manifest itself and its provoking “agents”, it is possible to assess whether this discrimination can be extended to ridesharing digital services.

In the literature, we find two types of discrimination: **direct** and **statistically or proxy**. According to [Brown 2019], **direct discrimination** is carried out by an “agent” based on observable personal characteristics of the person who suffered prejudice and causes a negative effect [Murphy 2002]. These personal characteristics can be race, gender, and sexual orientation, among others.

Statistically or proxy discrimination can occur consciously or unconsciously and is carried out when observable personal characteristics are used to infer about unobservable measures [Brown 2019] [Dovidio 2000]. Also, this kind of discrimination it's known as a **belief-based bias** [Monachou and Ashlagi 2019]. For example, when we have a service denied for a young person just because statistically we know that younger people have a lower income than people over 30 years old.

Another key concept found in the literature is the **taste-based** as one of a potential source of discrimination. According to [Monachou and Ashlagi 2019] the taste-based bias occurs when a person is not aware of his own prejudices and is associated with the absence of information of a person leading to discrimination. This form of bias is particularly insidious, as it operates beneath an individual's levels of conscious perception, transforming into a subtle yet powerful form of discrimination.

Now that we know the main concepts related to prejudice or bias, it is important to understand what it is **digital discrimination** or **discrimination in the online environment** occurs when a service is denied to a person or a group of individuals using his personal characteristics available on the services platform that can be used to identify and distinguish them such as symbols, colors, images, text, or graphics [Abramova 2020].

In the methodology section, we use a term called **Red Line** as a classification category to represent areas or neighborhoods that have high rates of violence or crime. In the next section, we adopt an **ontology** which is a semantic data structure that captures the relationships and concepts underlying a specific domain.

3. Ontology

The present work aims to understand and analyze whether there are signs of digital discrimination in the context of the ride-sharing service available in the city of Rio de Janeiro and whether this reason is related to some type of prejudice. In this context, the creation of an ontology dedicated to understanding the state of races, including cancellation, in order to identify whether the reason was due to prejudice is essential for a more in-depth and effective analysis of this problem.

An ontology is a formal and explicit representation of concepts and relationships in a domain of interest. Our ontology aims to map the main elements related to the status of the trip, whether there is discriminatory behavior involved or identified in cancellations, and the factors that lead to this behavior, as well as who practiced it (drivers or passengers). It provides a systematic and structured approach to understanding and addressing digital discrimination in ridesharing services. To better understand our domain, we analyzed application service information available on the Internet for driver terms and FAQs for passengers and drivers.

A user will download the application available on Android and iOS platforms. The versions of the two operating systems may vary or present differences in some features. After downloading and registering on the application, the user requests a ride by identifying their origin address (boarding) and entering their destination address (disembarkation), their payment method, and selecting the desired fare. After entering this information, the ride-sharing platform will search for the nearest drivers. Drivers have a specific application, where they register and, according to the platform, the login in the system is made to be in online mode to receive ride requests. After starting the system and being ready, he is capable of receiving ride requests, with initial entry information, such as the boarding region.

After the ride is accepted by a driver, the platform shares both passenger and driver information on their respective applications, for example, the ridesharing platform provides information to the passenger application about the driver who accepted the ride,

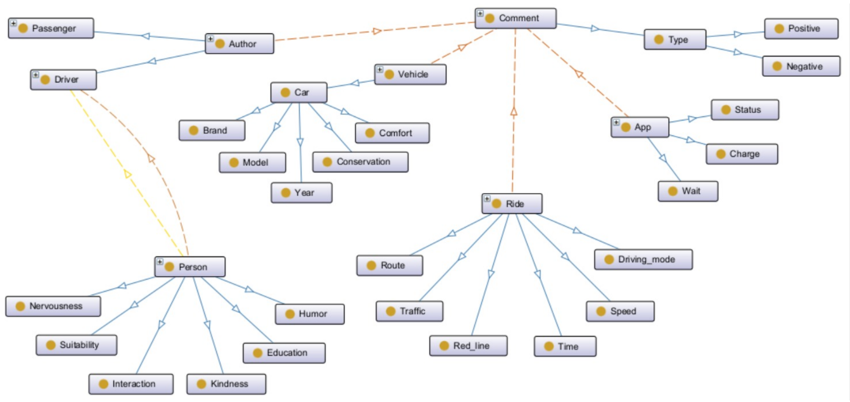

Figure 1. Ridesharing comments ontology

such as the name of the driver and information about the vehicle. After that, the ride can be canceled by both. If there is no cancellation, the passenger is boarded, so the driver starts the ride in his app until disembarking the ride, where the status changes to complete, and then moves on to the billing stage, where depending on the user's selected option, it can be done directly in the app or to the driver in the form of credit, debit or cash.

After billing, the process moves on to the evaluation stage. This is carried out through a rating system of 1 to 5 stars, where 1 signifies a poor experience and 5 denotes an excellent one. Additionally, there is an option to leave a comment, which can be used to express compliments or report any issues that may have occurred during the journey. In cases of great dissatisfaction with a platform's service, Brazilian users tend to adopt a complaints platform called Reclame Aqui, where users register their complaints and the platform may or may not provide feedback on the reported complaints. These comments can vary into two types, commendation, a positive type, and complaints, a negative type, regarding the provision of the service which is made up of the driver, his vehicle, the condition of the service provision and the functioning of the platform application itself. Comments about the driver can vary in relation to different characteristics, such as their driving mode, their education and attitudes, such as rudeness or kindness, to behaviors that should be banned in society, such as prejudice and harassment. Comments regarding the vehicle can be very diverse in relation to the vendor, model and age of the car to its condition and comfort. Other comments that can be found are in relation to the conditions for providing the service, including considering the route the driver took, the traffic encountered and the operation of the application itself, such as difficulties in registering, crashes or other difficulties in use.

4. Methodology

In our work, we selected an article that we can use as a baseline for our study and that can be reproduced with the data that can represent a ridesharing service used in Rio de Janeiro city. The research methodology adopted by [Miroslav Tushev and Mahmoud 2020] was to analyze the online feedback from the actors involved in the ridesharing service (drivers and passengers). The paper adopted the social network Twitter to represent this online feedback. Due to a particular characteristic of the Brazilian population, our proposal is to use the Reclame Aqui platform to obtain this feedback online.

4.1. Dataset

We extracted 210 complaints from user's platform on the Reclame Aqui website through a Python algorithm, using the BeautifulSoap and Selenium Webdriver libraries. The anonymous comments were saved in an Excel spreadsheet locally, where comments from rides with a finalized status application were included. The sample of these comments was anonymous, where there was only information about the comment and the score with which the ride was evaluated.

This dataset then had a total of 960 comments, in which pre-processing of the data was then carried out, such as removing lines that were brought with the phrase "Optional Comment", where the user did not make any comments in the application, just inserted an evaluation on the scoring system. After removing these lines, the final dataset resulted in 433 comments, where special characters generated, for example, by keyboard support configured on the smartphone, were removed.

4.2. Classification

After creating the dataset and pre-processing, we read and analyzed the 433 comments in pairs, where we manually classified each one into categories as shown in Table 1.

Of the 433 comments, 163 are multilabel and received more than one classification, as they contained complaints from 2 or 3 categories, totalizing 630 comments.

4.3. Data analysis

Despite the small sample of comments obtained, it was possible to identify the practice of discrimination, with a percentage of 1.9 percent, as shown in the graph in Figure 2. In this classification, we consider discrimination in relation to gender, including LGBT, ethnicity, ageism, weight, politics, and religion.

Of this percentage, 25 percent of users who suffered discrimination were female passengers, 8.3 percent were elderly passengers, and we also found cases of harassment of women. All religious and political discrimination was practiced by passengers in a percentage of 41.7 in relation to the total number of comments found with evidence of discrimination as we can see in Figure 3.

Two indicators suggest a more in-depth analysis, as it was not possible to identify whether there was direct or statistical discrimination by class or ethnicity. The largest of them, with 20.8 percent of complaints, were related to charging, where the passengers complained about drivers who canceled the ride or did not want to apply the discount selected by the passenger when requesting the ride. The second one, with 1.3 percent of complaints was categorized as Red Line, that is, where the destination address is located in communities or their surroundings. This indicator may be associated with public safety issues but also with discrimination by class or ethnicity.

5. Study results and discussion

Digital discrimination in shared services has been addressed from different aspects, but the main one found in the literature was diagnosis as in [Abramova 2020] [Brown 2019] [Ge 2018]. But approaches were also found from the aspect of the information system where the interest was divided into identifying biases in the algorithms adopted

Aggression;	Application;
Harassment;	Suitability;
Register;	Driver dissatisfaction;
Charge;	Cleaning;
Positive comments;	Driving mode;
Discrimination;	Red line;
Education;	Ride status; and
Conservation state;	Route.

Table 1. Classification categories

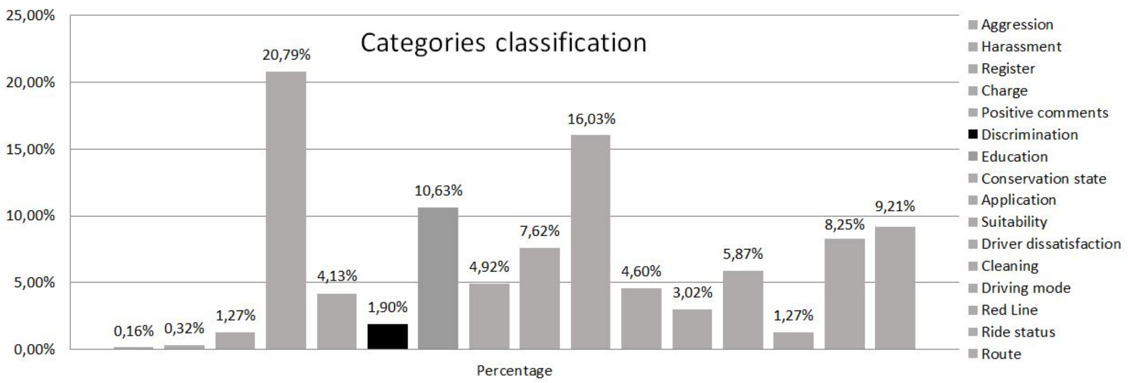

Figure 2. Discrimination evidence

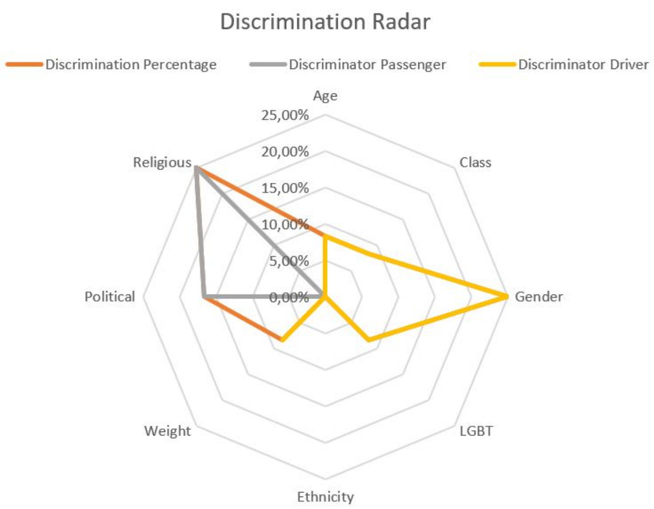

Figure 3. Discrimination radar

by shared ride platforms as in [Pandey and Caliskan 2021] and identifying biases and discrimination manifested by application drivers as in [Jorge Mejia 2020], by passengers as in [Alex Rosenblat and Hwang 2017] and by biases expressed by both as in [Miroslav Tushev and Mahmoud 2020].

Shared ride service platforms, with the aim of reducing discrimination on the part of drivers, began sending the least amount of information about the passenger to drivers when distributing the ride request. However, it was still possible to observe discrimination shortly after the acceptance and sharing of passenger characteristics such as name, gender, score, origin, and destination addresses [Jorge Mejia 2020].

Another point addressed in [Pandey and Caliskan 2021] was the bias in the algorithm of shared ride systems, where the price of ride fares varies not only with demand, but also with their location of origin or destination, where positive relationships were found, that is, it was identified that locations with a higher rate of acceptance of rides had higher fares, but also negative relationships, where locations with a higher rate of white population have lower rates while locations with a higher rate of non-white population have higher rates. In [Jorge Mejia 2020], the fare value is identified as one of the points to reduce the cancellation rate due to discrimination. The study suggests making this cost explicit as an attempt to reduce biased behavior. An interesting mitigation action found in the literature is a ride distribution model based on learning the history of acceptance and cancellation, that is, promoting a pairing of passengers and drivers, not only based on a scoring system but also with the identification of a bias in these platform users and with this it would be possible to expose them, according to [Monachou and Ashlagi 2019].

Comparing the results found in the literature with the data analyzed in our study, we suggest some points for discussion. The ride-sharing services available in the city of Rio de Janeiro present several differences from platform to platform. Based on our study, 7.4 percent of the complaints found were in relation to discounts not being applied or rides canceled due to the choice of a discounted ride. This difference suggests that users have different perceptions if this mitigating action works as expected or the dynamic system applied by this specific platform service, as we found in [Jorge Mejia 2020] and [Pandey and Caliskan 2021].

Another interesting point, which at the same time corroborates the analysis with the studies found in the literature, was in relation to the provision of passenger information, such as address of origin and destination, to the driver only after accepting the trip as [Miroslav Tushev and Mahmoud 2021], citeGe2016, [Brown 2019] and [Abramova 2020]. However, this was a point where we found opinions differing from the user platform. Passenger information is made available after acceptance of the trip as a way of mitigating discrimination, however, in the discrimination bar it is possible to identify that cancellation still occurs after this information is made available, and when not, passengers report that the service provided is impacted, causing embarrassment, discomfort and insecurity to the passenger who disembarks outside the location requested in the application.

6. Conclusion

In this study, it was possible to analyze that the main problem of this research is a topic of great relevance to society and there are opportunities to address it in the information system in order to promote mechanisms that reduce discrimination of any type, be it racial, gender, sexual orientation, religious or political association, of way to eradicate this behavior that is harmful to society. Our study, combined with an exploratory analysis of the state of the art in literature, proposed to answer the following questions:

- RQ1: Is there evidence of digital discrimination in the ridesharing application used in Rio de Janeiro city? Based on our analysis, it was possible to conclude that we found signs of prejudice, however, to answer this question, further exploration with more data will be necessary.
- RQ2: Is it possible to identify the factors that lead to discrimination? It was possible to identify that there are factors that can be associated with prejudice in particular towards women, with the comments, it was possible to identify that the majority of drivers are men, we found only 4 comments with reference to a driver woman, representing 0.63 percent, and 50 percent with positive comments.
- RQ3: What are the key concepts related to detecting Digital Discrimination in a ridesharing service? These concepts were identified in our analysis of the domain, where we proposed an ontology about it.

As [Miroslav Tushev and Mahmoud 2020], one limitation faced was the amount of data, plus the absence of user information for analysis, as all comments do not contain information and personal characteristics, it does not allow for a more in-depth analysis of some indicators that may or may not be related to discrimination due to prejudice, but it was not evident.

For future work, we propose to enlarge our dataset to include the complaints from Reclame Aqui website for more ridesharing services offered in Brazil and also propose a Machine Learning model to automatically classify those comments. Also, we can explore the dataset to evaluate and compare results on the detection of discrimination between different cities in Brazil, and if the differences between ridesharing platforms can increase or decrease the practice of discrimination.

References

Abramova, O. (2020). No matter what the name, we're all the same?

Alex Rosenblat, Karen E.C. Levy, S. B. and Hwang, T. (2017). Discriminating tastes: Uber's customer ratings as vehicles for workplace discrimination.

Batty, M. e. a. (2012). Smart cities of the future.

Brown, A. E. (2019). Prevalence and mechanisms of discrimination: Evidence from the ride-hail and taxi industries.

Dovidio, J. F. e. a. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level.

Ge, Y. e. a. (2018). Racial discrimination in transportation network companies.

Jorge Mejia, C. P. (2020). When transparency fails: Bias and financial incentives in ridesharing platforms.

Miroslav Tushev, F. E. and Mahmoud, A. (2020). Digital discrimination in sharing economy.

Miroslav Tushev, F. E. and Mahmoud, A. (2021). A systematic literature review of anti-discrimination design strategies in the digital sharing economy.

Monachou, F. G. and Ashlagi, I. (2019). Discrimination in online markets: Effects of social bias on learning from reviews and policy design.

Murphy, S. A. (2002). Appendix b: Audit studies and the assessment of discrimination.

Pandey, A. and Caliskan, A. (2021). Disparate impact of artificial intelligence bias in ridehailing.