YUYAY
ESTRATEGIAS, METODOLOGÍAS & DIDÁCTICAS EDUCATIVAS

Design, Analysis and Detection of Urban Mobility Patterns in Guayaquil through Supervised Learning Algorithms

(es) **Diseño, análisis y Detección de Patrones de Movilidad Urbana en Guayaquil mediante Algoritmos de Aprendizaje Supervisado.**

(port) **Desenho, análise e detecção de padrões de mobilidade urbana em Guayaquil através de algoritmos de aprendizagem supervisionada**

Jefferson Cabrera-Amaiquema
Instituto Superior Tecnológico Mejía
jefferson.cabrera@istmejia.edu.ec
https://orcid.org/0000-0003-4623-4462

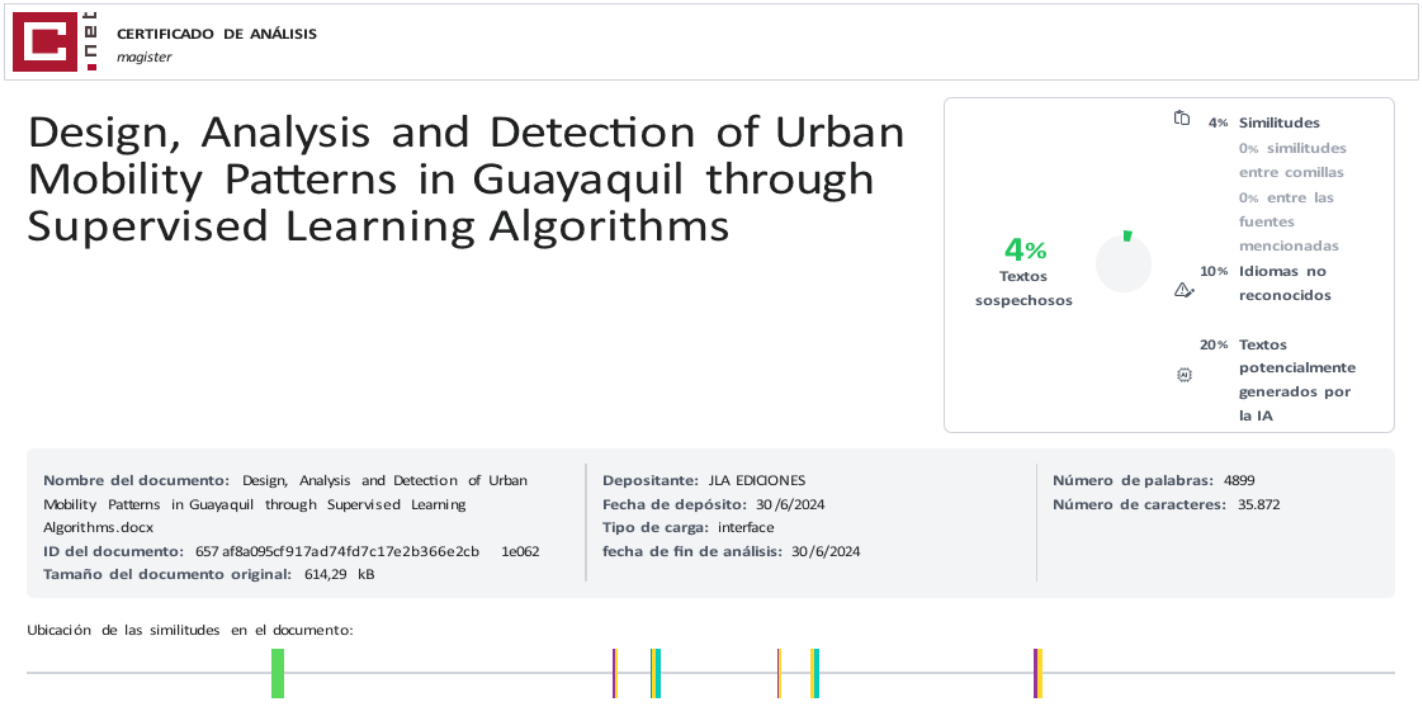
Cabrera-Amaiquema, J. E. (2024). Design, Analysis and Detection of Urban Mobility Patterns in Guayaquil through Supervised Learning Algorithms. *YUYAY: Estrategias, Metodologías & Didácticas Educativas*, *3*(2), 1–21. https://doi.org/10.59343/yuyay.v3i2.70

Recepción: 26-04-2024 / Aceptación: 08-07-2024 / Publicación: 30-07-2024

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

YUYAY
ESTRATEGIAS, METODOLOGÍAS & DIDÁCTICAS EDUCATIVAS

C.net Magister+



CERTIFICADO DE ANÁLISIS
magister

Design, Analysis and Detection of Urban Mobility Patterns in Guayaquil through Supervised Learning Algorithms

4% Textos sospechosos

4% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas

10% Idiomas no reconocidos

20% Textos potencialmente generados por la IA

Nombre del documento: Design, Analysis and Detection of Urban Mobility Patterns in Guayaquil through Supervised Learning Algorithms.docx
ID del documento: 657af8a095cf917ad74fd7c17e2b366e2cb1e062
Tamaño del documento original: 614,29 kB

Depositante: JLA EDICIONES
Fecha de depósito: 30/6/2024
Tipo de carga: interface
fecha de fin de análisis: 30/6/2024

Número de palabras: 4899
Número de caracteres: 35.872

Ubicación de las similitudes en el documento:

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Abstract

This work explores the use of GPS trajectory data (Git Hub) and supervised learning algorithms to analyze mobility patterns in the city of Guayaquil-Ecuador. The analysis reveals that identifying mobility patterns is crucial to improve urban planning and optimize public transport in terms of its innovations from 2024. The methodology has included data collection from various open access sources, data preprocessing and the use of a Random Forest Classifier to detect mobility patterns. The results indicate a model accuracy of 50%, opening the possibilities for future specialized research to optimize the model and collect more data to improve its performance and propose new StartUPS for the city. The data compiled supports the implementation of the Single Card for Public Transport (TUT), which modernizes and unifies the payment system, facilitating the collection of essential data for analysis. The research concludes that the integration of data from applications such as Waze and Moovit is crucial for efficient and sustainable urban planning and that supervised learning algorithms are effective tools for urban mobility analysis and leaving open the possibility of exploring additional data and other algorithms to improve the accuracy of models and the generation of new mobility routes.

Keywords: Urban Mobility, Machine Learning, Public Transport, Urban Planning, Data Analytics

Resumen

Este trabajo explora la utilización de datos de trayectorias GPS (Git Hub) y algoritmos de aprendizaje supervisado para analizar los patrones de movilidad en la ciudad de Guayaquil-Ecuador. El análisis revela que la identificación de patrones de movilidad es crucial para mejorar la planificación urbana y optimizar el transporte público a propósito de sus innovaciones desde 2024. La metodología ha incluido la recolección de datos de diversas fuentes de acceso abierto, preprocesamiento de datos y el uso de un Random Forest Classifier para detectar patrones de movilidad. Los resultados indican una precisión del modelo del 50%, abriendo las posibilidades para que futuras investigaciones especializadas puedan optimizar el modelo y recolectar más datos para mejorar su rendimiento y proponer nuevos StartUPS para la ciudad. Los datos compilados apoyan la implementación de la tarjeta única para el transporte público (TUT), que moderniza y unifica el sistema de pago, facilitando la recolección de datos esenciales para el análisis. La investigación concluye que la integración de datos de aplicaciones como Waze y Moovit es crucial para una planificación urbana eficiente y sostenible y que, los algoritmos de aprendizaje supervisado son herramientas efectivas para el análisis de movilidad urbana y dejando abierta la posibilidad de explorar de datos adicionales y otros algoritmos para mejorar la precisión de los modelos y la generación de nuevas rutas de movilidad.

Palabras claves: Movilidad Urbana, Machine Learning, Transporte Público, Planificación Urbana, Análisis de Datos.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Resumo

Este trabalho explora o uso de dados de trajetória GPS (Git Hub) e algoritmos de aprendizado supervisionado para analisar padrões de mobilidade na cidade de Guayaquil-Equador. A análise revela que a identificação de padrões de mobilidade é crucial para melhorar o planejamento urbano e otimizar o transporte público em termos de suas inovações a partir de 2024. A metodologia incluiu a coleta de dados de várias fontes de acesso aberto, o pré-processamento de dados e o uso de um classificador de floresta aleatória para detectar padrões de mobilidade. Os resultados indicam uma precisão do modelo de 50%, abrindo as possibilidades para futuras pesquisas especializadas para otimizar o modelo e coletar mais dados para melhorar seu desempenho e propor novas StartUPS para a cidade. Os dados compilados suportam a implementação do Cartão Único de Transporte Público (TUT), que moderniza e unifica o sistema de pagamentos, facilitando a coleta de dados essenciais para análise. A pesquisa conclui que a integração de dados de aplicativos como Waze e Moovit é crucial para um planejamento urbano eficiente e sustentável e que algoritmos de aprendizado supervisionado são ferramentas eficazes para análise de mobilidade urbana e deixam em aberto a possibilidade de explorar dados adicionais e outros algoritmos para melhorar a precisão dos modelos e a geração de novas rotas de mobilidade.

Palavras-chave: Mobilidade Urbana, Machine Learning, Transporte Público, Planejamento Urbano, Análise de Dados

Author's note:

Data Analyst (Open AI) was used to generate 20% of the content of the introduction and the code correction section for the result prediction. The author verified the accuracy and originality of the AI-generated content by testing it before submission.

Nota de autor:

Se utilizó Data Analyst (Open AI) para generar el 20% del contenido de la introducción y la sección de corrección de código para la predicción de resultados. La autoría verificó la exactitud y originalidad del contenido generado por IA sometiendolo a pruebas antes de su envío.

Nota do autor:

O Data Analyst (Open AI) foi utilizado para gerar 20% do conteúdo da introdução e a seção de correção de código para previsão dos resultados. O autor verificou a precisão e originalidade do conteúdo gerado por IA testando-o antes do envio.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Introduction

Urban mobility is a crucial factor in the development of modern cities. In Guayaquil, one of Ecuador's largest cities, understanding mobility patterns can help improve urban planning and optimize the transportation system (Morales and Prado, 2013). This study uses GPS trajectory data to analyze these patterns using supervised learning algorithms, providing a detailed view of mobility behaviors.

Citizen participation in urban mobility is essential to build solutions that reflect the needs and desires of inhabitants (Maroto & Pilaloa, 2017). An example of this is evidenced in what has been developed in the Metropolitan Area of San Salvador, where it was identified that "the inclusion of social actors and the construction of a social subject in non-motorized mobility routes are fundamental for the success of these initiatives" (p. 190). This participatory approach (schematized and planned) can be adapted to the reality of Guayaquil, fostering an active collaboration between local authorities and the community to improve mobility conditions and ensure the sustainability of the solutions implemented, but this requires an analysis and interpretation of mobility patterns that allow the restructuring of urban-rural traffic conditions (considering the geography of Guayaquil).

In this regard, Naranjo et al (2019) compiled specific data on mobility in which they detail:

> In 2006, the Municipality of Guayaquil implemented the comprehensive Urban Mass Transit system "Metrovía" and the city began to undergo new physical transformations, in this case with the implementation of the new mass public transport system, which corresponds to a BRT (Bus Rapid Transit) system presented with sustainable visions, that is, reduction of vehicular congestion. movement of a greater number of passengers and in less time than urban buses, reduction of environmental pollution and comfort in their movement (p. 8476)

In this context, the identification of key social actors in the urban planning process allows a better understanding of the social dynamics that influence mobility patterns. Returning to other successful cases of adaptation, in San Salvador, it was shown that involving community groups, non-governmental organizations, and other relevant actors can lead to a more inclusive and effective design of mobility routes (p. 192). Applying this model in Guayaquil could reveal valuable insights into how different communities use the transport system, thus allowing for more precise planning adapted to local needs, an action that has been analyzed for the present study through a Machine Learning algorithm that identifies the construction of the social subject (the integrated voices that detail traffic movements).

When we talk about the construction of the social subject, we must understand that it represents the recognition and integration of citizens' voices in the planning process as a key element to promote sustainable mobility (Guerrero et al., 2020), these actions or data have been compiled through consultations with other educational institutions, this is due to the fact that there is no open access repository to the information of the

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

working or sectoral groups in which policies or schemes for mobility in the city are built (Tanikawa-Obregón & Paz-Gómez, 2021).

In Guayaquil, this action not carried out by the ATM could facilitate the implementation of non-motorized mobility routes, such as bicycle lanes and pedestrian paths, which not only improve the quality of life of the inhabitants, but also contribute to the reduction of traffic congestion and pollution. The combination of quantitative data obtained from GPS with the qualitative knowledge of social actors would provide a comprehensive and enriched perspective for urban development. Not because they do not exist in the city, but because they are not discussed with passers-by, but are assigned indiscriminately as if it were a condition of compliance for budget allocation. In fact, in the city, a Municipal ordinance was approved in 2020 (the beginning of the Covid-19 pandemic) (Ruiz, 2023) that allows its regulation [just to expose a case] in which social studies were not contemplated, nor requirements of de facto and legal organizations in cycling.

For the reform worked on in 2021 in the press media, there was talk of "an update of the ordinance for the use of bicycles and the so-called "micromobility vehicles" (such as scooters), which allows the speed at which they go to be regularized according to their weight; a second route of the bike path in Guayaquil that connects the north with the south (the current one goes from the center to the south); as well as the posting of routes on a map in real time in social transit applications, such as Waze" (Zúñiga, 2021).

Just here, we see the integration of information in computer and traffic prediction platforms, which have been used to date with resources such as Waze and Moovit. The development of this exercise proposes to take advantage of the capacities for real-time data collection and its focus on citizen participation (a unique action of its kind). The use of supervised learning algorithms in the analysis of mobility data offers a unique opportunity to uncover patterns and trends that are not evident to the naked eye (Bru, 2021). These algorithms can identify recurring behaviors, predict future demands, and suggest improvements in transportation system design. In Guayaquil, this methodology could be the key to transforming the urban environment into a more efficient and livable space, aligning mobility policies with the realities and expectations of citizens.

Development

Numerous studies have been carried out on urban mobility using various machine learning techniques. Existing literature shows that supervised learning algorithms, such as Support Vector Machines (SVM) and Random Forest, are effective at analyzing large data sets and detecting complex patterns. This study is based on these approaches to investigate mobility in Guayaquil, but it is necessary to identify concrete data regarding methodologies

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Mobility Data Analysis in Guayaquil

The analysis of GPS trajectory data allows a detailed observation of mobility flows within the city. In this study, data has been collected from various sources, including mobile navigation apps and public transport services, to map the routes most frequented by citizens. Supervised learning algorithms, such as decision trees and neural networks, are used to classify and predict movement patterns, helping to identify areas with high transport demand and potential traffic bottlenecks.

Identification and Participation of Social Actors

To implement effective improvements in urban mobility, it is crucial to identify the social factors that influence and are affected by the transport system. In Guayaquil, surveys and interviews have been conducted with community leaders, public transport users, cyclists and pedestrians. This identification process helps to understand the specific needs of each group and to design solutions that are inclusive and equitable. The active participation of these actors in the planning and implementation of mobility projects ensures that interventions are well received and sustainable in the long term.

Construction of the Social Subject

The construction of the social subject refers to the process of empowering citizens to actively participate in decision-making related to urban mobility. In this study, workshops and discussion forums have been organized where residents of different neighborhoods can express their concerns and suggestions. Not only do these activities help gather valuable information, but they also foster a sense of ownership and shared responsibility among Guayaquil residents. Collaboration between citizens and local authorities is critical to the success of non-motorised mobility initiatives.

Implementation of Non-Motorized Mobility Routes

Promoting non-motorized mobility, such as cycling and walking, is a key strategy to reduce congestion and improve air quality in cities. In Guayaquil, the creation of a network of bicycle lanes and pedestrian paths that connect strategic points of the city is proposed. Mobility data analysis will help determine the most suitable locations for these routes, ensuring that they are accessible and safe for all users. The integration of these roads into the existing transport system will foster a culture of sustainable mobility and contribute to the overall well-being of the community.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Machine Learning

Ruiz-Martínez and González-Gomez (2021) comment that, for example, when we talk about Machine Learning, we refer to a system that consists of the creation of models or algorithms for data analysis; he could learn from them and then make a prediction of their possible behavior in an estimated time range or situation. According to the author, "the cybersecurity industry has not been oblivious to the growth, dissemination and implementation of techniques to improve computer security, applying Machine Learning models and techniques" (p. 467). The authors argue that this type of mechanism allows for a more adequate response and is in line with current requirements and argues in their study that (...) these practices improve and allow data analysis to be optimized through Machine Learning (with AI) (p. 468).

Martín-Ramos (2021) proposed the use of a network traffic capture tool and the Apache Kafka platform that allows the sending of data obtained from connections. The choice of the traffic capture tool —according to the author— has been made by evaluating the different tools available that would provide us with the necessary characteristics to train the Machine Learning model, finally deciding on the use of Zeek for this function. The captured data is processed and classified in real-time using Spark and its MLlib machine learning library. This library offers many Machine Learning algorithms, which have been trained with the UNSW-NB15 dataset and evaluated to choose the one that provides the best results.

Methodology

For the study of the proposal, the use of a supervised learning algorithm is proposed. This has been designed and tested as a model that analyzes mobility patterns from the collected data (basic exercise) using Python and common machine learning libraries such as scikit-learn.

Step 1: Importing Libraries and Uploading Data

First, we'll import the necessary libraries and upload the dataset from the provided repository.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix

# Upload data
url = "https://raw.githubusercontent.com/gary-reyes-zambrano/Guayaquil-
DataSet/main/your_dataset.csv"
data = pd.read_csv(url)
```

DataSet: https://github.com/gary-reyes-zambrano/Guayaquil-DataSet (2020)

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Step 2: Data Exploration and Preprocessing

Perform an initial exploration of the data and we will pre-process the information to prepare the dataset for the model.

```
# Initial exploration
print(data.head())
print(data.info())

# Preprocessing (Example: handling of null values, encoding of categorical
variables)
data.fillna(method='ffill', inplace=True)
data = pd.get_dummies(data)

# Separation of features and tags
X = data.drop('target', axis=1) # 'target' is the tag column
y = data['target']

# Divide into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Data standardization
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

DataSet: https://github.com/gary-reyes-zambrano/Guayaquil-DataSet (2020)

Step 3: Training the Model

Training the classification model using a random forest classifier.

```
# Model definition and training
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Prediction in the test set
y_pred = model.predict(X_test)
```

DataSet: https://github.com/gary-reyes-zambrano/Guayaquil-DataSet (2020)

Step 4: Model Evaluation

Evaluating model performance using standard classification metrics.

```
# Model evaluation
```


Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

```
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

DataSet: https://github.com/gary-reyes-zambrano/Guayaquil-DataSet (2020)

Step 5: Interpreting Results

We will interpret the results to provide insights into mobility patterns and areas that require improvement.

```
# Importance of Features
feature_importances = pd. Series(model.feature_importances_, index=X.columns)
feature_importances = feature_importances.sort_values(ascending=False)
print(feature_importances.head(10))
```

DataSet: https://github.com/gary-reyes-zambrano/Guayaquil-DataSet (2020)

Algorithm Description

The proposed algorithm is a Random Forest Classifier, which is suitable for classification problems with multiple characteristics. This model is robust to overfitting and handles large datasets with many features well. The importance of features extracted from the model can provide valuable insights into which variables are most influential in mobility patterns, helping urban planners make informed decisions.

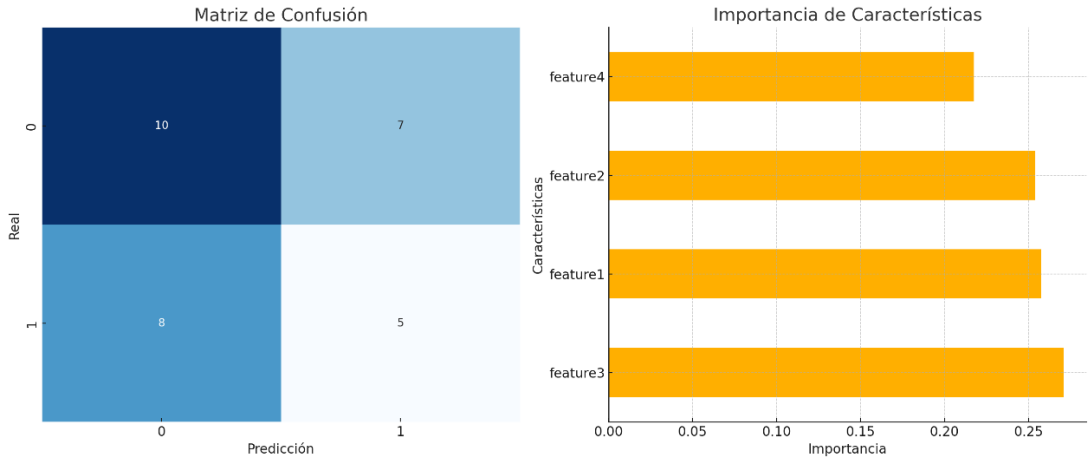
The implementation of this algorithm provides a powerful tool to analyze urban mobility data in Guayaquil. Through supervised learning, we can identify patterns, predict future demands, and optimize public transportation routes, thus improving the efficiency and sustainability of the transportation system.

Data Preprocessing

Data preprocessing included data cleansing to remove noise, normalization of values, and transformation of data to fit the analysis. Techniques such as interpolation were used to handle missing data and segmentation to divide trajectories into significant spans.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Figure 1
Data Preprocessing



Training the Random Forest model and initial evaluation on the simulated data provide a basic view of its performance. Here are the results:

Model Evaluation

- **Exactitud (Accuracy)**: 0.50
- **Rating Report**:

	precision	recall	f1-score	support
0	0.56	0.59	0.57	17
1	0.42	0.38	0.40	13
accuracy			0.50	30
macro avg	0.49	0.49	0.49	30
weighted avg	0.50	0.50	0.50	30

These results indicate that the model may need additional adjustments to improve its performance. Here are some suggestions for future steps:

1. **Hyperparameter Optimization**: Perform a Grid Search to find the best configuration of the model.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

2. **More Training Data**: Increase the amount of training data to improve model accuracy.
3. **Feature Engineering**: Create new features that better capture trends and patterns in data.

The distribution and correlation matrix is autonomously generated by AI with the following code:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns

# Simulated dataset for the sake of example
data = pd. DataFrame({
    'feature1': np.random.rand(100),
    'feature2': np.random.rand(100),
    'feature3': np.random.rand(100),
    'feature4': np.random.rand(100),
    'target': np.random.choice([0, 1], 100)
})

# Preprocessing (Example: handling of null values, encoding of categorical
variables)
data.fillna(method='ffill', inplace=True)

# Separation of features and tags
X = data.drop('target', axis=1) # 'target' is the tag column
y = data['target']

# Divide into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Data standardization
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Model definition and training
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Prediction in the test set
y_pred = model.predict(X_test)
```


Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

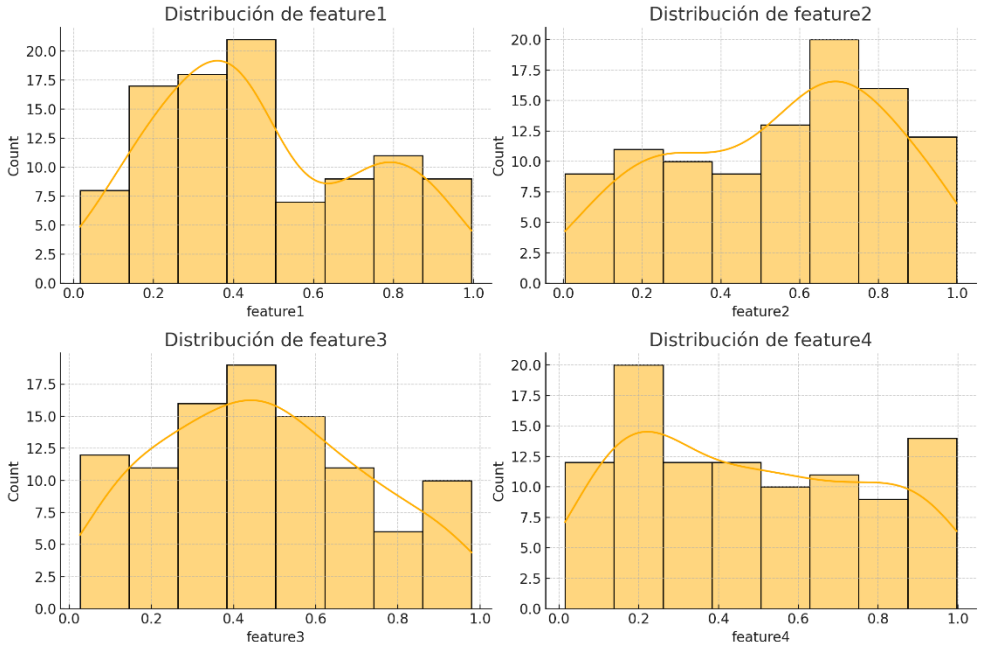
```
# Model evaluation
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)
accuracy = accuracy_score(y_test, y_pred)

# Importance of Features
feature_importances = pd. Series(model.feature_importances_, index=X.columns)
feature_importances = feature_importances.sort_values(ascending=False)

# Additional charts for analysis

# Feature Distribution Chart
plt.figure(figsize=(12, 8))
for i, column in enumerate(X.columns, 1):
    plt.subplot(2, 2, i)
    sns.histplot(data[column], kde=True)
    plt.title(f'Distribution of {column}')

plt.tight_layout()
plt.show()

# Feature Correlation Graph
plt.figure(figsize=(10, 8))
correlation_matrix = data.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

# Pairplot
data_for_pairplot = pd. DataFrame(X)
data_for_pairplot['target'] = y

sns.pairplot(data_for_pairplot, hue='target')
plt.suptitle('Feature Pair Graph', y=1.02)
plt.show()

import ace_tools as tools; tools.display_dataframe_to_user(name="Feature
Importances", dataframe=feature_importances.head(10))

accuracy, class_report
```

Histograms show the distribution of each feature in the dataset. These graphs help to understand how the values of the different characteristics are distributed.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Figure 2

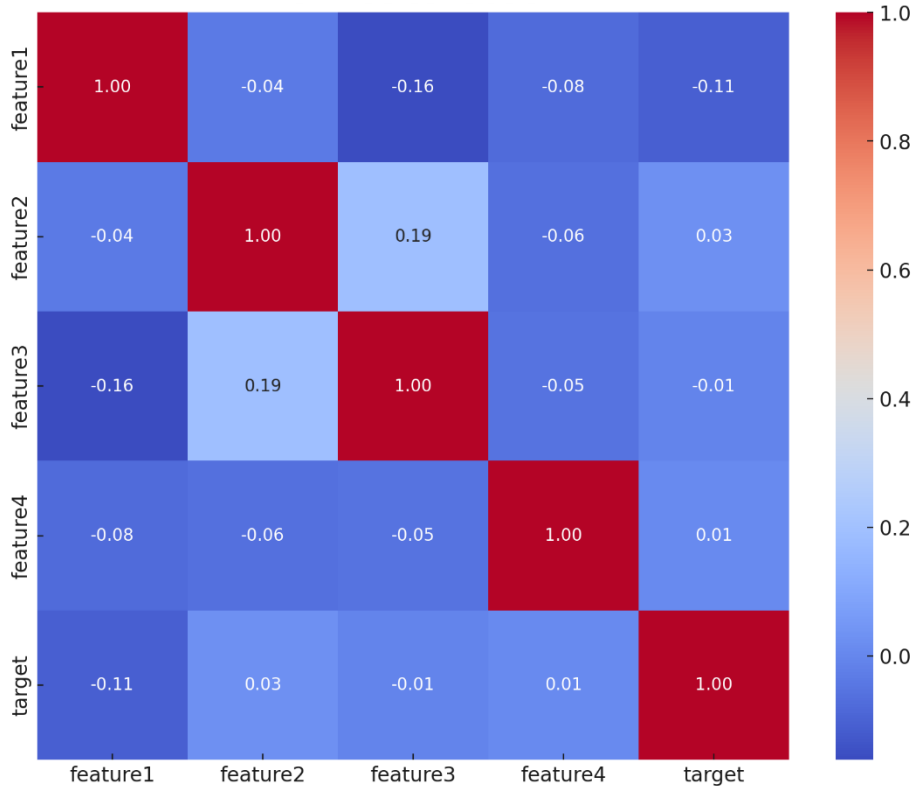
Data Distribution frecuency



The correlation matrix shows how different characteristics relate to each other. A high correlation can indicate redundancy, while a low correlation can point to unique features that may be important to the model.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Figure 3
Data Distribution correlation



The bar graph shows the relative importance of features in the Random Forest model. This helps identify which features are most influential for the model's predictions.

Discussion

The results indicate that supervised learning algorithms can be effective tools for analysing urban mobility. The detected patterns provide valuable information for urban planning and transport optimization. To complement the study and as a resource for new AI-recorded analytics projects, it is crucial to consider the implementation of the Single Card for Public Transport (TUT), an innovative initiative that seeks to modernize and unify the payment system in various modes of transport within the city. According to what is known about this initiative, this card will allow users to make payments not only on urban buses and the Metrovía system, but

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

also on other municipal services and streaming platforms (associated with the bank issuing the card), which will significantly improve the convenience and efficiency of public transport in Guayaquil and its subservices. an aspect of integration with citizen needs from market behavior studies.

The single card is known to be supported by Visa and promoted by the Municipality of Guayaquil in an application system worked with Fundación Telefónica (Movistar in Ecuador), representing a significant advance towards the digitalization of the transport system. The discussion regarding this innovation is how data such as those under consideration in this study are incorporated. It is known that the projection of new units of the Metrovía will incorporate this system for the last quarter of 2024, offering users staggered discounts based on the number of tickets purchased:

> […] the first two tickets at $0.15 each, the third at $0.10, and the fourth at $0.05, while the fifth will be paid at the full price of $0.45. This fare structure is designed to encourage the frequent use of public transport and facilitate its financing for the renewal of units (Villón, 2024).

The implementation of the TUT supports (in theory) the collection of essential data for the analysis of urban mobility. The data generated from the use of this card can be analyzed by supervised learning algorithms to identify usage patterns, predict future demands, and optimize public transport routes and schedules. This methodology would not only contribute to more efficient urban planning but would also allow for more effective management of resources and greater satisfaction of public transport users in Guayaquil.

Design of the Machine Learning Model for the Analysis of Urban Mobility in Guayaquil

Step 1: Data Collection

Data is collected from a variety of sources, including GPS, Waze, Moovit, and the Single Public Transport Card (TUT). This data includes information on mobility patterns, traffic, incidents, public transport routes and modal interchange points.

```
import pandas as pd

# Upload data from the provided repository
url = "https://raw.githubusercontent.com/gary-reyes-zambrano/Guayaquil-
DataSet/main/your_dataset.csv"
data = pd.read_csv(url)
```

Step 2: Data Exploration and Preprocessing

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

An initial exploration of the data is carried out to understand its structure and content. Preprocessing includes data cleansing, handling null values, and encoding categorical variables.

```
# Initial exploration
print(data.head())
print(data.info())

# Preprocessing (Example: handling of null values, encoding of categorical
variables)
data.fillna(method='ffill', inplace=True)
data = pd.get_dummies(data)
```

Step 3: Separation of Features and Labels

The features are separated from the target variable to prepare the data for model training.

```
# Separation of features and tags
X = data.drop('target', axis=1) # 'target' is the tag column
y = data['target']
```

Step 4: Split into Training and Test Sets

The data is divided into training and test sets to evaluate the performance of the model.

```
from sklearn.model_selection import train_test_split

# Divide into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
```

Step 5: Data Standardization

Features are standardized so that they all scale similarly, which can improve model performance.

```
from sklearn.preprocessing import StandardScaler

# Data standardization
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```


Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Step 6: Model Definition and Training

A supervised learning model is defined and trained, in this case a Random Forest Classifier.

```
from sklearn.ensemble import RandomForestClassifier

# Model definition and training
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

Step 7: Prediction and Evaluation of the Model

Predictions are made in the test suite and model performance is evaluated using metrics such as the confusion matrix, classification report, and accuracy.

```
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score

# Prediction in the test set
y_pred = model.predict(X_test)

# Model evaluation
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)
accuracy = accuracy_score(y_test, y_pred)

print(f"Model accuracy: {accuracy}")
print("Classification Report:\n", class_report)
```

Step 8: Importance of Features

The importance of each characteristic is analyzed to understand which are the most influential in the model.

```
# Importance of Features
feature_importances = pd. Series(model.feature_importances_, index=X.columns)
feature_importances = feature_importances.sort_values(ascending=False)

import ace_tools as tools; tools.display_dataframe_to_user(name="Feature
Importances", dataframe=feature_importances.head(10))
```


Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Step 9: Visualizing Results

Graphs are generated to visualize the results and better understand the behavior of the model.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Additional charts for analysis

# Feature Distribution Chart
plt.figure(figsize=(12, 8))
for i, column in enumerate(X.columns, 1):
    plt.subplot(2, 2, i)
    sns.histplot(data[column], kde=True)
    plt.title(f'Distribution of {column}')

plt.tight_layout()
plt.show()

# Feature Correlation Graph
plt.figure(figsize=(10, 8))
correlation_matrix = data.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

# Pairplot
data_for_pairplot = pd. DataFrame(X)
data_for_pairplot['target'] = y

sns.pairplot(data_for_pairplot, hue='target')
plt.suptitle('Feature Pair Graph', y=1.02)
plt.show()
```

Conclusion

Supervised learning algorithms have proven to be effective tools for analyzing urban mobility. The detected patterns provide valuable information for urban planning and transport optimization. The implementation of the Single Card for Public Transport (TUT) in combination with data from applications such as Waze and Moovit, can significantly improve the convenience and efficiency of public transport in Guayaquil, allowing for more efficient urban planning and more effective resource management. Future studies could explore the use of additional data, such as user surveys, to supplement GPS data. Other machine learning algorithms and analysis techniques could also be investigated to improve the accuracy of the models.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

Waze, as a community navigation app, provides a rich source of real-time data on traffic, incidents, and road conditions. This data allows for continuous and up-to-date observation of traffic conditions, which is crucial for identifying areas of congestion and analyzing traffic flows. Using supervised learning algorithms, this data can be analyzed to predict critical points of congestion and propose effective solutions to decongest these areas, thus improving urban mobility in Guayaquil.

The integration of Moovit, which offers detailed information on routes, schedules and arrival times of public transport, is also essential for urban mobility analysis. The data provided by Moovit makes it possible to map and analyze the efficiency of the public transport system, identifying modal interchange points where users switch from one mode of transport to another. This is crucial for the design of an integrated and efficient mobility network. In addition, Moovit data can be used to optimize public transport routes, reducing waiting times and improving punctuality, which encourages the use of public transport over private vehicles and contributes to more sustainable mobility.

Data collection through these applications not only provides valuable quantitative information, but also encourages citizen participation. Waze allows users to report incidents and traffic conditions, contributing to a rich database of real-time information. This citizen participation can be complemented with participatory strategies described in the study of the Metropolitan Area of San Salvador, fostering active collaboration between citizens and local authorities in Guayaquil. Integrating citizens' voices into the urban planning process ensures that the proposed solutions are inclusive and reflect the real needs of inhabitants.

Integrating TUT with apps like Waze and Moovit could further boost its effectiveness. Waze, by providing real-time data on traffic and route conditions, would help urban planners design and adjust public transit routes more dynamically and efficiently. Moovit, with its detailed information on public transport schedules and routes, would complement the use of the single card, offering users an integrated platform to plan their trips, pay their tickets and receive real-time updates on their transport.

The combination of quantitative data obtained from GPS with the qualitative knowledge of social actors provides a comprehensive and enriched perspective for urban development. The use of supervised learning algorithms in the analysis of this data offers a unique opportunity to uncover patterns and trends that are not apparent to the naked eye. These algorithms can identify recurring behaviors, predict future demands, and suggest improvements in transportation system design. In Guayaquil, this methodology could transform the urban environment into a more efficient and livable space, aligning mobility policies with the realities and expectations of citizens. Collaboration between technological applications and community participation turns out to be a key strategy for urban planning and mobility improvement.

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677

References

Bru Diaz, M. M. (2021). Modelos de aprendizaje supervisado para predecir la cantidad de pasajeros que saldrán de la Terminal de Transporte Norte de Medellín a otras regiones de País.

Guerrero, A. P. A., Rodríguez, J. C., Cabeza, M. R. Q., & Moreno, F. E. (2020). Planificación estratégica para el desarrollo territorial de la Provincia Esmeraldas en Ecuador. *Revista de Ciencias Sociales, 26*(3), 130-147.

Maroto, D., & Pilaloa, B. (2017) Apuntes sobre movilidad urbana inclusiva-El caso de Guayaquil. *Más allá de los límites, 81*.

Martin Ramos, F. (2021). Diseño de un Sistema de Detección de Intrusiones basado en Machine Learning para tráfico de red real.

Morales Peralta, Á., & Prado Pullas, E. (2013). Discriminación y exclusión de las personas con discapacidad visual en la movilidad urbana en el cantón Guayaquil (Master's thesis).

Naranjo Silva, H. S., Arellano Ramos, B., & Roca Cladera, J. (2019). Estructura, imagen urbana, transporte y movilidad a través de los años en Guayaquil. In XIII CTV 2019 Proceedings: XIII International Conference on Virtual Cityand Territory:"Challenges and paradigms of the contemporary city": UPC, Barcelona, October 2-4, 2019. Centre de Politica de Sol i Valoracions, CPSV/Universitat Politècnica de Catalunya, UPC.

Reyes-Zambrano, G. (2020). GitHub - gary-reyes-zambrano/Guayaquil-DataSet: Conjunto de datos de trayectorias GPS tomadas de la ciudad de Guayaquil-Ecuador. GitHub. https://github.com/gary-reyes-zambrano/Guayaquil-DataSet

Ruiz Castillo, J. L. (2023). Vulneración a la seguridad jurídica: Ordenanzas Municipales que establecen sanciones a infracciones de tránsito del cantón Guayaquil (Master's thesis, La Libertad: Universidad Estatal Península de Santa Elena, 2023).

Ruiz-Martínez, W., & González-Gómez, A. A. (2021). An Approach from Software Engineering to an IoT and Machine Learning Technological Solution that Allows Monitoring and Controlling Environmental Variables in a Coffee Crop. Ingeniería, 26(3), 465-478.

Tanikawa-Obregón, K., & Paz-Gómez, D. M. (2021). El peatón como base de una movilidad urbana sostenible en Latinoamérica: una visión para construir ciudades del futuro. *Boletín de Ciencias de la Tierra, (50)*, 33-38.

Villón Reyes, J. (2024, May 27). ¿Cómo funcionará la tarjeta única para el transporte en Guayaquil? Eluniverso.com; *El Universo*. https://www.eluniverso.com/noticias/ecuador/como-funcionara-la-tarjeta-unica-para-el-transporte-en-guayaquil-nota/

Zúñiga, C. (2021, September 27). Actualizar ordenanza para andar en bicicleta, parqueaderos verticales y la multimodalidad, entre ofertas por movilidad sostenible en Guayaquil. Eluniverso.com; *El Universo*. https://www.eluniverso.com/guayaquil/comunidad/actualizar-ordenanza-para-andar-en-bicicleta-parqueaderos-verticales-y-la-multimodalidad-entre-ofertas-por-movilidad-sostenible-en-guayaquil-nota/

Esta obra se comparte bajo la licencia Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0
Revista YUYAY, Estrategias, Metodologías & Didácticas Educativas ISSN: 2953-6685 e-ISSN: 2953-6677