Cruz et al. *EPJ Data Science* (2025) 14:3
https://doi.org/10.1140/epjds/s13688-024-00515-9

EPJ Data Science
a SpringerOpen Journal

EPJ.org

RESEARCH Open Access

A new approach to estimate neighborhood socioeconomic status using supermarket transactions and GNNs

Eduardo Cruz[1*], Monica Villavicencio[1], Carmen Vaca[1], Lisette Espín-Noboa[2,3] and Nervo Verdezoto[4]

*Correspondence:
escruz@espol.edu.ec
[1]Escuela Superior Politecnica del Litoral, Guayaquil, Ecuador
Full list of author information is available at the end of the article

Abstract

Ending poverty in all its forms everywhere remains the number one Sustainable Development Goal of the United Nations 2030 Agenda. Governments face challenges in measuring socioeconomic status with fine spatial resolution because traditional data collection methods, such as censuses and surveys, are time-consuming, labor-intensive, performed at long intervals, and cover only a limited population. This work is a data-driven study to analyze the digital traces left by humans in supermarket transactions and model the relationship between consumption behavior and the average per capita income, proposing a proxy to estimate socioeconomic status at the urban neighborhood level. We analyze more than 20 million supermarket shopping transactions in Guayaquil, the most populated city in Ecuador. Using customer consumption data, we created a basket graph and fed it into a graph neural network to predict neighborhood socioeconomic status. The model was trained with spectral and spatial convolutional filters using cross-validation to select the best approach for the prediction. The results show that the Chebyshev spectral convolutional filter has the highest predictive power to predict the socioeconomic status of the neighborhood, with $R^2 = 0.91$. Our proposed approach contributes to measuring socioeconomic status at the neighborhood level to support policymakers in making informed decisions about resource allocation according to the needs of different geographical areas.

Keywords: Neighborhood socioeconomic status; Item embedding; Basket graph; Graph neural network; Spectral convolutional filter; Per capita income

1 Introduction

In 2015, the United Nations (UN) adopted the Agenda for Sustainable Development as a universal call to action to end poverty, protect the planet, and ensure that all people can enjoy a more sustainable planet by 2030. This agenda comprises seventeen Sustainable Development Goals (SDGs), which urge all countries to take immediate action [1]. SDG 1 aims to end poverty in all its forms everywhere, which remains one of the most pressing issues in the world [2]. Despite efforts to alleviate poverty, projections indicate that more than 600 million will still live in extreme poverty in 2030, struggling to access basic ne-

Springer

cessities such as healthcare, education, water, and sanitation [3]. To monitor these trends and anticipate future challenges, governments must periodically assess the socioeconomic status (SES) of the population at a high level of spatial granularity. This approach helps to evaluate economic and social positioning relative to other populations and supports data-driven decisions on resource allocation based on regional needs.

Governments employ SES measurement as an analytical tool to assess individuals' income, education, occupation, and living conditions within a population. National statistical institutes (NSIs) are responsible for collecting, measuring, and updating SES data. They employ traditional data collection methods such as censuses and surveys to measure SES, including monetary unidimensional indicators like per capita income (PCI) and non-monetary multidimensional indices like the index of multiple deprivation (IMD). However, censuses are time-consuming and labor-intensive, leading NSIs to conduct them over extended intervals, generally every ten years. This infrequent update delays the availability of comprehensive wealth distribution data required for governments to develop timely economic recovery programs [4–6]. To address the limitations of censuses, NSIs often perform surveys focused on population samples and conducted at shorter intervals. Although surveys help to update SES data more frequently, they generally report results with coarse spatial granularity, usually at administrative levels 1 and 2 (province and city) [7, 8]. When governments need SES insights with a finer spatial resolution, such as urban neighborhoods, to understand wealth distribution and design resource allocation programs, surveys only cover some of the urban neighborhoods within a city for the limited population coverage.

In response to these challenges, we propose a data-driven study that analyzes digital traces left by individuals in supermarket transactions to model the relationship between consumption behavior and socioeconomic status with a finer spatial resolution than the city level, specifically the urban neighborhood level. Our study is located in Guayaquil, the most populated city in Ecuador, and uses a digital dataset of more than 20 million transactions from a supermarket chain to estimate neighborhood socioeconomic status, represented by average PCI. First, we model consumption behavior through a graph structure that captures common item combinations purchased at supermarkets in urban neighborhoods. Then, we train a graph neural network (GNN) with spectral convolutional filters to learn how these consumption patterns relate to neighborhood socioeconomic status. Our contributions are summarized as follows:

- We characterize the consumption behavior from a graph perspective to identify neighborhood socioeconomic status.
- We built a GNN model as a new way to predict socioeconomic status at the urban neighborhood level.

2 Related work

An extensive body of research has studied surveys and diverse digital sources such as call detail records (CDR), satellite imagery, geographic information, social media, and grocery shopping data to extract features used in novel models to estimate socioeconomic status. Below, we summarize the latest efforts in using digital sources and constructing statistical and machine learning models to predict real-value socioeconomic indicators with fine spatial resolution.

The expansion of mobile phone use has provided a rich and broad source of information about the socioeconomic characteristics of regions. The CDRs obtained from mobile

network operators provide real-time data that can be used to model customer behavior in measuring socioeconomic indicators. Hernandez et al. [9] studied call activity recorded in network operator towers, extracting consumption (incoming and outgoing calls) and mobility features (how far and how often a customer travels). They designed Voronoi polygons to represent network cell coverage, which were merged with the geographical units where poverty rates were surveyed. The features based on activity were aggregated at the polygon network level using customers' home location, and a new set of features was included, such as the number of customers entering the network polygon and living outside of it and how often they visit it. Finally, the authors built a linear regression model to predict poverty rates at the municipal level in Guatemala with a prediction power of $R^2 = 0.76$. Similarly, Steele et al. [10] analyzed mobile phone usage variables such as incoming and outgoing calls, incoming and outgoing text, percent of nocturnal calls, number and entropy of places visited, radius of gyration, interactions per contact, entropy of contacts calculated per customer and aggregated at a granular spatial resolution using the Voronoi polygons with the tower geolocation. They built a Bayesian model to predict the wealth index (WI) extracted from the demographic and health survey (DHS) for Namibia, Nepal and Bangladesh. The model achieved a prediction power of $R^2 = 0.66$, $R^2 = 0.61$, and $R^2 = 0.64$, respectively.

Satellite imagery data sources have also been widely explored to study the relationship between spatial composition and poverty. Zhao et al. [11] extracted features such as the radiance value of nighttime lights, principal components of structure and texture, land cover, and accessibility from NPP-VIIRS NTL, Google satellite images, land cover maps and OpenStreetMap (OSM) road maps. They built a random forest regression model to estimate the household wealth index from Bangladesh and Nepal's Demographic and Health Surveys at a spatial area of 10 square kilometers around. The prediction power of the model was $R^2 = 0.70$ for Bangladesh and $R^2 = 0.61$ for Nepal. Similarly, Xu et al. [12] proposed an XGBoost model to predict the integrated poverty index (IPI) for an impoverished area in southwest China. The model used spatial variables such as nighttime light, traffic accessibility, altitude, flat area coverage, impervious surface coverage, cropland coverage, road density, and risk index extracted from multisource geospatial data (NPP-VIIRS NTL, SRTM DEM, FROM-GLC, OSM). The model achieved a precision of $R^2 = 0.68$ to predict the IPI at the village level. Likewise, Li et al. [13] extracted image characteristics from Google Earth street view and used a land cover dataset to build a support vector regression model to predict urban poverty at the community level for the Jiangxia and Huangpi districts of Wuhan. The proposed model reached a performance of $R^2 = 0.53$ for both districts.

The growth in the use of social networks among populations has generated interest in studying the relationship with socioeconomic indicators. Piaggesi et al. [14] examined the Facebook advertising audience to predict the socioeconomic conditions of residents in Atlanta, Santiago, Bogotá, and Casablanca. They performed queries using the Facebook marketing API into a geographic area of interest with a fixed radius of 1 km. For each query, they requested the number of users who matched more than 47 attributes associated with gender, marital status, education, travel, interests, religion, technology, and connectivity. They developed a linear regression model to predict median household income, socioeconomic strata, and multidimensional poverty rate. The model reached a $R^2 = 0.93$ for Santiago, while it was between $R^2 = 0.46$ and $R^2 = 0.56$ for the other three cities, with users

older than 25. Similarly, Fatehkia et al. [15] collected data on Facebook users aged 18+ in the Philippines and India, focusing on device types and network connections. These data points were then correlated with regional WI values derived from the DHS. The study used linear and tree-based regression models to predict WI using Facebook advertising data, population density, and regional indicators. The results showed that Facebook penetration and the type of device or network used were strong predictors of WI, achieving a cross-validated $R^2 = 0.608$ for the Philippines and $R^2 = 0.563$ for India.

Similarly to our approach, graphs are utilized as data structures to model relationships, and when combined with neural networks, they have been used to develop architectures capable of predicting poverty. Cao et al. [6] analyzed open geospatial data from multiple sources, including remote sensing imagery and OpenStreetMap road networks, to construct multi-view graphs that capture the correlations between regions, including region adjacency and region similarity graphs. They built a multi-view graph neural network (MVGNN) to predict the gross domestic product (GDP) at China's county level. The proposed model reached a performance of $R^2 = 0.82$.

Several studies have analyzed the relationship between supermarket shopping behavior and socioeconomic status at the individual and household level [16–18]. Pechei et al. [18] studied how SES, supermarket choice, and shopping behaviors influence the healthfulness of food purchases. They studied panel data from 24,879 UK households, focusing on fruits, vegetables and unhealthy food purchases. The households were stratified by SES, and their supermarket choices were classified into high, medium and low price tiers. The results revealed that households who shop primarily in low-price supermarkets purchased fewer fruits and vegetables and more unhealthy foods than those who shop in high-price supermarkets. Furthermore, in terms of shopping behavior, households visiting shorter trips were associated with unhealthy food choices, while those visiting a larger variety of store chains tended to purchase more fruits and vegetables. The study concluded that both supermarket choice and shopping behaviors affect the healthiness of purchases. Similarly, Vinkeles et al. [16] explored the relationship between SES and supermarket shopping behavior, focusing on purchasing unhealthy and energy-dense foods. They collected data from the shopping dockets of 204 shoppers (102 from low SES areas and 102 from high SES areas). They found that shoppers from low SES areas purchased significantly more non-core, unhealthy food items (especially chips and sugar-sweetened beverages) than those from high SES areas. The results suggest that socioeconomic factors play a significant role in food purchasing habits, with individuals from lower SES areas being at higher risk of poor dietary choices. However, these studies have employed traditional data collection methods, which have the limitations of being time-consuming and expensive tasks that cannot be carried out with high temporal frequency. On the other hand, similar to our work, a recent study has used supermarket transactions registered in digital data to predict socioeconomic status at a fine spatial resolution. Bannister et al. [19] analyzed grocery shopping data from Tesco stores to explore the relationship between shopping habits and IMD at London's lower super output area (LSOA) level. They built three random forest regression models with different combinations of features derived from grocery store data, such as the fraction of purchased products divided into seventeen categories and the weight of nutrients divided into seven categories. The best models were those trained with all the features and only with the features of the nutrients that reached a predictive

power of $R^2 = 0.57$ in the prediction of IMD, confirming the relationship between the food people buy in a supermarket and the level of LSOA deprivation in London.

Most studies have analyzed CDRs, satellite imagery, geographic information, and social media to provide valuable information to predict socioeconomic status, addressing the limitations of traditional data collection methods at fine spatial resolutions. However, with the availability of other digital data sources capturing the purchasing activities of customers in urban neighborhoods, there is significant potential to explore and leverage deep learning models to uncover non-linear relationships between purchasing behavior and SES at the neighborhood level. Such approaches could improve predictive power, offering a more effective means to measure the economic rhythms of the population.

3 Dataset

We describe alternative digital data sources used in this work to model the relationship between consumption behavior and socioeconomic status at a finer spatial resolution than the city level.

3.1 Urban neighborhoods

Ecuador is organized into administrative boundaries to facilitate public management in a decentralized or deconcentrated manner. The following administrative levels are available in digital format: level 1 (provinces), level 2 (cantons/cities), and level 3 (urban parishes). However, the boundaries at the third administrative level, which offer a finer spatial resolution than at the city level, contain multiple geographical divisions with distinct socioeconomic characteristics. Therefore, this study is conducted at the urban neighborhood level, which provides a more granular spatial resolution than the urban parish level and is defined as a residential geographic area where most households share similar socioeconomic characteristics.

However, the boundaries of urban neighborhoods in Guayaquil are not publicly available in digital format. To address this, we constructed a geographic dataset of urban neighborhood boundaries by redrawing polygons in QGIS, using sources of geographic information such as Google Maps, Nominatim with OpenStreetMap, and Wikimapia. The resulting dataset, in GeoJSON format, includes 97 neighborhoods with an area ranging from 2 to 9 km^2.

3.2 Supermarket transactions

Ecuador is home to three dominant supermarket chains with national coverage. These chains have strengthened their reach by creating various store formats to meet the needs of consumers across the socioeconomic spectrum, including those of lower income groups [20]. As a result, even the smallest markets in urban neighborhoods have lost market share due to the expansion of supermarket chains [21]. These chains offer various products, including groceries, personal hygiene products, and cleaning supplies.

One of these supermarket chains provided access to anonymized purchase data from its 257 stores. Our study focuses explicitly on the 53 stores in Guayaquil, which are geographically distributed among urban neighborhoods of varying socioeconomic status. Guayaquil has the highest concentration of chain stores.

In Guayaquil, the chain recorded more than 20 million shopping transactions from January 1, 2016, to December 31, 2018. Each transaction includes details such as the transac-

Table 1 Format of supermarket shopping transactions

Date	Transaction ID	Customer ID	Item ID	Qty	Store ID
2016-01-01	1001	101	101	10	1
2016-01-01	1001	101	202	5	1
2016-01-01	1002	902	402	8	2
2016-01-01	1002	902	602	3	2
.	.	.	.	.	.
.	.	.	.	.	.

Table 2 Format of geographic information for the stores

Store ID	Latitude	Longitude
1	−2.25754	−79.89032
2	−2.27512	−79.87929
.	.	.
.	.	.

Table 3 Format of the features for the items

Item ID	Item family	Unit price
101	CANNED FISH	2.45
202	MILK CARTON	1.10
402	SODA	0.56
602	SAUSAGES	0.47
.	.	.
.	.	.

tion date, transaction ID, customer ID, item ID, quantity, and store ID (refer to Table 1). Furthermore, the supermarket chain provided two complementary datasets:

1. A dataset linking each store with geographic coordinates (see Table 2).
2. A dataset classifying each item by family and price (refer to Table 3).

It is important to note that sensitive customer information was excluded from these datasets.

We defined the catchment area for each supermarket store as a 500-meter radius around its geolocation point. Next, we performed a spatial join between the catchment areas and the urban neighborhood polygons. The urban neighborhood with the largest intersecting area was assigned to each corresponding supermarket store. This process resulted in 53 stores in Guayaquil spatially linked to 36 urban neighborhoods.

3.3 Neighborhood socioeconomic status

The Ecuadorian national institute of statistics and censuses (INEC) has conducted the National survey of employment, unemployment, and underemployment (ENEMDU) annually since 1993. ENEMDU is a national quarterly survey that examines employment conditions, characteristics of the labor market, economic activities, and sources of income from the population. INEC employs a spatially disaggregated sampling frame that relies on census tracts, with each tract containing an average of 150 dwellings and an average land area of 200 m^2 per dwelling. Due to the high resource and time requirements to conduct sociodemographic research at the national level, the quarterly sample size is limited to 2438 census tracts selected by stratified random sampling in all provinces [22]. However, the surveyed census tracts only cover a portion of the city's urban neighborhoods,

Table 4 Summary of the surveyed urban neighborhoods from 2016 to 2018

Year	Surveyed neighborhoods					NSES			
	q1	q2	q3	q4	Total	Min	Max	Avg	Std
2016	26	25	27	26	104	145.21	567.27	254.45	115.50
2017	28	26	26	27	107	138.39	575.87	255.43	114.77
2018	24	26	27	26	103	134.76	600.35	257.49	117.01

making it challenging to analyze the socioeconomic realities of urban neighborhoods that were not included in the sample.

In Guayaquil, we collected socioeconomic survey data from 2016 to 2018, aligning with the period of collecting data from supermarket transactions. The geographic boundaries of the census tracts are publicly available in digital format from INEC. To assess the neighborhood socioeconomic status (NSES), we first performed a spatial join between the boundaries of the surveyed census tracts and the 36 urban neighborhoods that contain supermarkets. We then determined the average per capita income for each neighborhood, which serves as the ground-truth socioeconomic indicator. This process resulted in 314 neighborhood samples with both NSES surveyed and supermarket transaction data. Our supervised inference method uses these samples to estimate socioeconomic status (details are provided in Table 4).

4 Methodology

The proposed methodology is a data-driven process that transforms supermarket transactions into a graph representation for a deep learning model to predict NSES. The process involves three phases:

1. Creating an item embedding $\hat{I}$ based on their family and price.
2. Representing a basket graph G using transaction data from urban neighborhood stores to represent item co-purchases.
3. Constructing a GNN to predict NSES using the basket graph G.

4.1 Item embedding

We develop a module to encode item features as a fixed-size embedding used as node attributes in the graph representation. Given $F = \{f_1, f_2, \ldots, f_i\}$ representing the set of item families, a Doc2Vec model is trained [23] to encode each item family in a real-valued vector of size 8 using the set of item families as the corpus, producing a new set of item family embeddings $\hat{F} = \{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_i\}$. Additionally, each item family $f_i \in \hat{F}$ has a set of prices $P_{fi} = \{p_1, p_2, \ldots, p_j\}$. The prices are standardized using MinMaxScaler with a range between 0.1 and 0.9 to accommodate the totals where the price is multiplied by quantity. The process produces a new set of standardized item prices $\bar{P}_{fi} = \{\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_j\}$ for each item family f_i. Therefore, an item embedding combines the family embedding and the standardized price, producing a real-valued set of item embeddings $\hat{I} = \{(\hat{f}_i, \bar{p}_j) \mid \hat{f}_i \in \hat{F}, \bar{p}_j \in \bar{P}_{fi}\}$, where $\hat{I} \in \mathbb{R}^d$, and $d = 9$.

4.2 Basket graph

We convert supermarket transactions into a graph representation called basket graph [24] to model relationships between purchased items and understand the consumption behavior of customers within an urban neighborhood. Given $T_n(r) = \{t_1, t_2, \ldots, t_m\}$ as the

Table 5 Summary of the nodes (V) and links (E) statistics for the created basket graphs

Year	Basket graphs	V				E			
		Min	Max	Avg	Std	Min	Max	Avg	Std
2016	104	1005	1642	1372	127	72,680	287,510	171,594	49,280
2017	107	1145	1675	1437	116	99,746	296,652	179,885	45,725
2018	103	1018	1604	1262	116	73,454	258,756	138,079	40,622

set of supermarket transactions for the neighborhood n during the quarter r; where $t_m = \{((\hat{f}_1, \bar{p}_1), q_1), ((\hat{f}_2, \bar{p}_1), q_2), \ldots, ((\hat{f}_i, \bar{p}_j), q_k)\}$ is defined as the set of item embeddings $(\hat{f}_i, \bar{p}_j) \in \hat{I}$ with their quantities $q_1, q_2, \ldots, q_k$ purchased in the transaction t_m. The basket graph is defined as a weighted attributed graph $G = (V, E, W)$, where $V \in \mathbb{R}^{|V| \times d}$ are the attributed nodes that belong to the set of item embeddings $\hat{I}$ purchased in supermarkets of the urban neighborhood n, $E \in \mathbb{R}^{|E|}$ denotes the set of links representing pairs of item embeddings purchased within the same transaction t_k, and $W \in \mathbb{R}^{|E| \times 1}$ represents the weight set for each link in E. The weight $w_{xy} \in W$ for the link between the item embedding x and y is calculated using the following equation:

$$w_{xy} = \frac{1}{S} \sum_{s=1}^{S} (\bar{p}_x * q_x) + (\bar{p}_y * q_y) \tag{1}$$

S represents the number of transactions where the item embeddings x and y were purchased together, $\bar{p}$ is the standardized price, and q is the purchased quantity. Links with the number of transactions S below the third quartile are removed to exclude random item co-purchases, reducing the density of the graph and highlighting the most common item co-purchases that best characterize the urban neighborhood n. This process resulted in 314 basket graphs representing customer consumption behavior in urban neighborhoods during all quarters from 2016 to 2018 (details are provided in Table 5). The 211 basket graphs from 2016 to 2017 are samples to train the model, and the 103 basket graphs from 2018 are the samples to test the model.

4.3 Graph neural network model

Our focus is on creating a GNN model to learn about the consumption behavior of the customers represented by the basket graph structure to predict NSES. The model architecture incorporates three convolutional layers of 32 channels with a nonlinear activation function (Softmax). Each convolutional layer extracts features from the previous layer's outputs, and with the nonlinear activation function, the information generated by the convolution layer is transmitted to the next layer. In addition, a dropout $p = 0.25$ is applied to the result of each convolutional layer to avoid overfitting. Finally, an average mean pooling as an aggregation function (AGG) is used to produce a fixed-size representation of the entire graph connected to a linear layer to construct the predicted $\mathbf{N\bar{S}ES}$.

GNN layers can be constructed with two types of convolutional filters. The spatial approach provides filter localization with a finite size, operating on the local neighborhood of nodes to understand the properties of a node based on its local neighbors. In contrast, the spectral approach performs an eigendecomposition of the Laplacian matrix to capture the underlying structure of the graph [25]. A five-fold cross-validation training process is performed using the 211 basket graphs from 2016 to 2017 as input, with the mean square

Table 6 Results of five-fold cross-validation for spectral and spatial convolutional filters

Fold #	ChebConv		SAGEConv		GraphConv		GCNConv	
	RMSE	R2	RMSE	R2	RMSE	R2	RMSE	R2
Fold 1	39.11	0.91	60.43	0.78	72.39	0.69	72.22	0.69
Fold 2	34.15	0.91	64.45	0.66	61.28	0.70	37.66	0.88
Fold 3	46.68	0.82	50.90	0.78	79.85	0.46	44.85	0.82
Fold 4	49.53	0.74	56.95	0.65	60.54	0.62	68.48	0.50
Fold 5	35.55	0.92	56.87	0.78	68.82	0.69	73.21	0.64
Avg	**41.01**	**0.86**	57.93	0.74	68.57	0.63	59.29	0.71
Std	6.09	0.07	4.48	0.06	7.21	0.09	14.98	0.13

error (MSE) as the learning objective, since the NSES is a continuous value. This setup is used to select the optimal convolutional filter approach between spectral (ChebConv) and spatial (SAGEConv, GraphConv, GCNConv) for maximizing predictive power.

The predictive power of the model is evaluated using two performance metrics: the coefficient of determination (R^2) to evaluate the goodness of fit and the root mean square error (RMSE) to calculate the error (results are presented in Table 6). The results indicate that the GNN model with the Chebyshev spectral convolutional filter achieved the highest predictive power, with the highest $R^2 = 0.86$ and lowest $RMSE = 41.01$ on average, demonstrating a good balance between bias and variance.

Therefore, the proposed GNN model is implemented with the spectral approach for its convolutional layers using Chebyshev filters of size 4 ($K = 4$). Each layer in this setup utilizes the following equation:

$$\mathbf{X}' = \sigma\left(\sum_{k=1}^{K} \mathbf{Z}^{(k)} \cdot \mathbf{\Theta}^{(k)}\right) \tag{2}$$

$$\begin{aligned} \mathbf{Z}^{(1)} &= \mathbf{X} \\ \mathbf{Z}^{(2)} &= \hat{\mathbf{L}} \cdot \mathbf{X} \\ \mathbf{Z}^{(k)} &= 2 \cdot \hat{\mathbf{L}} \cdot \mathbf{Z}^{(k-1)} - \mathbf{Z}^{(k-2)} \end{aligned} \tag{3}$$

where σ is a nonlinear function (Softmax), K is the Chebyshev filter size; $\mathbf{Z}^{(k)}$ is the Chebyshev polynomial of order k computed recursively and evaluated in the normalized Laplacian $\hat{\mathbf{L}}$ (a diagonal matrix of scaled eigenvalues) $\hat{\mathbf{L}} = \frac{2\mathbf{L}}{\boldsymbol{\lambda}\mathbf{max}} - \mathbf{I}$; and $\mathbf{\Theta}^{(k)}$ is a vector of learnable polynomial coefficients [25] (details are shown in Fig. 1).

5 Results

5.1 GNN model evaluation

The proposed GNN model was evaluated using basket graphs from urban neighborhoods surveyed during the quarters of 2018. This evaluation assessed the model's predictive power against a test set that was not used during training. Two performance indicators, R^2 and $RMSE$ were used to measure the accuracy of the model (residuals are shown in Table 7 and the results are presented in Table 8). To analyze the result, urban neighborhoods were classified into four groups based on their quartile distribution: low, medium-low, medium-high and high NSES. The results show that the highest error occurred for the NSES of *Bellavista* in all quarters, while high errors for *Guayacanes* and *Garzota* were

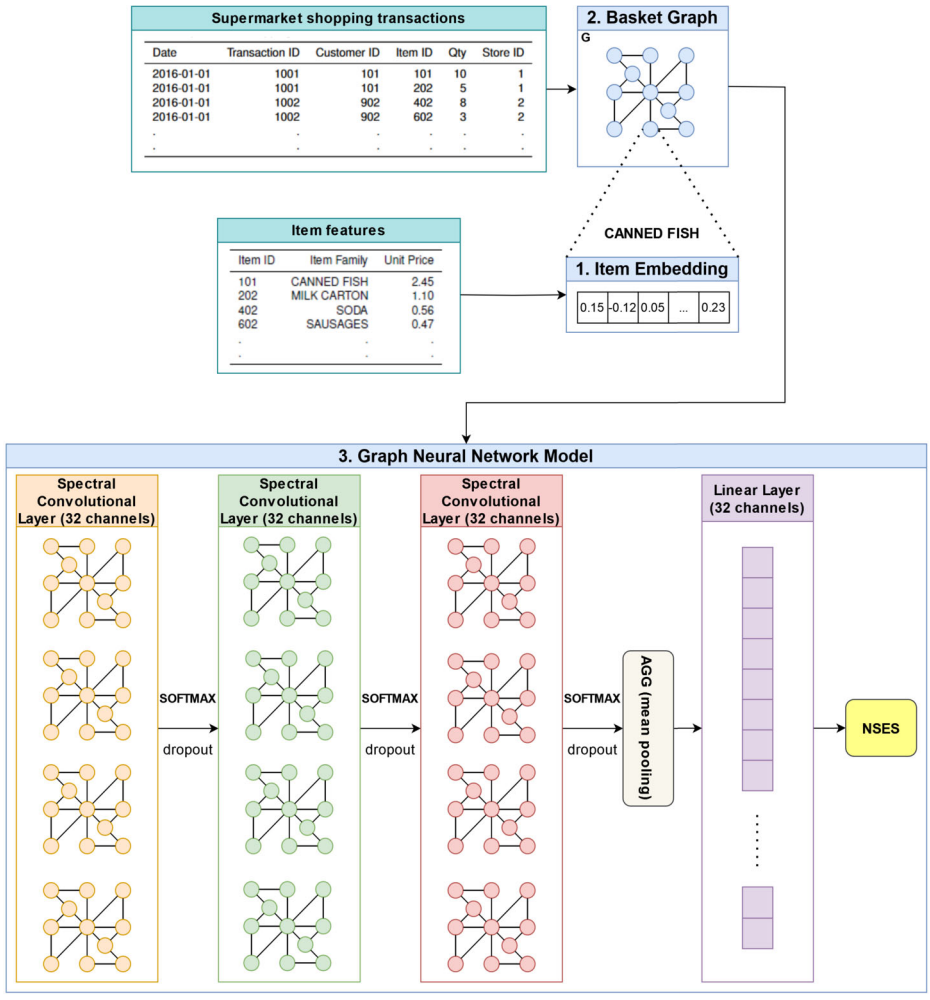

Figure 1 Architecture of the proposed GNN model

observed only in the first and second quarters, respectively. *Bellavista* and *Garzota* are urban neighborhoods with high NSES, while *Guayacanes* has a medium-high NSES.

The high prediction error for neighborhoods with high NSES could be attributed to the limited number of samples that represent that socioeconomic level during training. In contrast, the model exhibited the lowest prediction errors for neighborhoods with medium-high, medium-low, and low NSES. Consequently, the GNN model predicts NSES with an accuracy of up to 91% for Guayaquil's urban neighborhoods in 2018, with a margin of error of ±33.70 USD (see Table 8). These results confirm the GNN model's strong predictive capability for NSES in urban neighborhoods, even for quarters not included in the model's training.

The correlation between the NSES predicted by the GNN model and the NSES surveyed by the INEC was calculated for urban neighborhoods in all quarters of 2018. The correlation analysis revealed a very strong positive correlation, ($\rho = 0.96$), with statistical significance ($p < 0.001$) (see Fig. 2). The correlation holds a very strong positive correlation when the top 10% of the urban neighborhoods surveyed are discarded, ($\rho = 0.91$), with statistical significance ($p < 0.001$). This confirms the ability of the GNN model, us-

Table 7 Residuals between predicted and surveyed NSES values

Neighborhood	Surv.	Pred.	Error	Neighborhood	Surv.	Pred.	Error
Quarter 1				Quarter 2			
Bellavista	**589.71**	**496.22**	**93.49**	***Bellavista***	**594.88**	**489.55**	**105.33**
Garzota	516.74	493.42	23.32	***Garzota***	**521.91**	**442.81**	**79.10**
Samanes	476.48	483.86	–7.39	Samanes	481.65	458.75	22.90
Alborada	415.12	399.39	15.73	Guayacanes	382.98	383.39	–0.42
Guayacanes	**377.81**	**473.89**	**–96.08**	Sauces	374.66	321.87	52.78
Mucho Lote	241.91	243.83	–1.91	Garcia Moreno	270.17	273.90	–3.73
Floresta	237.76	173.84	63.93	Mucho Lote	247.08	226.90	20.18
Mapasingue	225.22	219.30	5.92	Floresta	242.93	185.04	57.89
Febres Cordero	223.15	194.32	28.83	Mapasingue	230.39	214.58	15.80
Letamendi	213.03	186.18	26.85	Febres Cordero	228.32	208.24	20.07
Guasmo Norte	204.42	174.63	29.79	Letamendi	218.20	265.13	–46.93
Crt. Consuelo	204.13	205.18	–1.05	Juan Montalvo	211.07	186.30	24.76
Abel Gilbert	198.97	211.28	–12.32	Guasmo Norte	209.59	183.01	26.58
Guasmo Sur	196.49	191.03	5.45	Crt. Consuelo	209.30	201.15	8.16
Guasmo Oeste	196.13	189.70	6.43	Abel Gilbert	204.14	210.82	–6.68
Btll. Suburbio	194.67	184.39	10.28	Guasmo Sur	201.66	216.88	–15.22
Vergeles	192.31	182.63	9.69	Guasmo Oeste	201.30	195.58	5.72
Puerto Lisa	192.25	202.32	–10.06	Btll. Suburbio	199.84	200.86	–1.02
Pascuales	186.31	189.63	–3.32	Vergeles	197.48	184.21	13.28
Isla Trinitaria	184.59	182.06	2.53	Puerto Lisa	197.42	184.64	12.78
Bastion Popular	165.52	172.29	–6.76	Pascuales	191.48	182.75	8.73
El Fortin	157.29	174.40	–17.11	Isla Trinitaria	189.76	231.15	–41.39
Nv. Prosperina	139.17	175.98	–36.82	Bastion Popular	170.69	196.40	–25.71
Flor de Bastion	134.76	178.73	–43.97	El Fortin	162.46	178.23	–15.77
				Nv. Prosperina	144.34	179.74	–35.40
				Flor de Bastion	139.93	176.10	–36.17
Quarter 3				Quarter 4			
Bellavista	**598.05**	**501.59**	**96.46**	***Bellavista***	**600.36**	**495.25**	**105.11**
Garzota	525.08	497.50	27.59	Garzota	527.39	498.47	28.92
Samanes	484.82	462.12	22.69	Alborada	425.77	427.26	–1.49
Alborada	423.46	472.30	–48.83	Guayacanes	388.46	399.44	–10.98
Guayacanes	386.15	406.53	–20.38	Sauces	380.14	320.34	59.79
Sauces	377.83	335.69	42.14	Garcia Moreno	275.65	271.28	4.37
Garcia Moreno	273.34	254.29	19.05	Mucho Lote	252.56	259.42	–6.85
Mucho Lote	250.25	199.58	50.67	Floresta	248.41	230.63	17.78
Floresta	246.10	238.75	7.35	Mapasingue	235.87	195.72	40.14
Mapasingue	233.56	218.77	14.79	Febres Cordero	233.80	208.01	25.79
Febres Cordero	231.49	187.14	44.35	Letamendi	223.68	207.82	15.86
Letamendi	221.37	239.51	–18.14	Juan Montalvo	216.55	214.59	1.95
Juan Montalvo	214.24	199.74	14.49	Guasmo Norte	215.07	182.27	32.79
Guasmo Norte	212.76	179.84	32.92	Crt. Consuelo	214.78	209.53	5.25
Crt. Consuelo	212.47	187.65	24.82	Abel Gilbert	209.62	229.02	–19.40
Abel Gilbert	207.31	241.90	–34.59	Guasmo Sur	207.14	188.94	18.20
Guasmo Sur	204.83	187.04	17.79	Guasmo Oeste	206.78	193.18	13.60
Guasmo Oeste	204.47	190.13	14.34	Btll. Suburbio	205.32	201.90	3.42
Btll. Suburbio	203.01	219.09	–16.08	Vergeles	202.96	189.65	13.31
Vergeles	200.65	185.19	15.46	Puerto Lisa	202.90	188.29	14.62
Puerto Lisa	200.59	189.38	11.22	Pascuales	196.96	187.32	9.64
Pascuales	194.65	200.52	–5.87	Isla Trinitaria	195.24	192.29	2.95
Isla Trinitaria	192.93	220.88	–27.95	Bastion Popular	176.17	182.62	–6.45
Bastion Popular	173.86	182.05	–8.19	El Fortin	167.94	183.16	–15.22
El Fortin	165.63	179.58	–13.94	Nv. Prosperina	149.82	183.45	–33.63
Nv. Prosperina	147.51	184.73	–37.22	Flor de Bastion	145.41	188.25	–42.84
Flor de Bastion	143.10	178.98	–35.88				

Table 8 Summary of GNN models' performance

Quarter	R^2	*RMSE*
1	0.91	35.09
2	0.89	36.47
3	0.92	32.57
4	0.92	30.53
All year	0.91	33.70

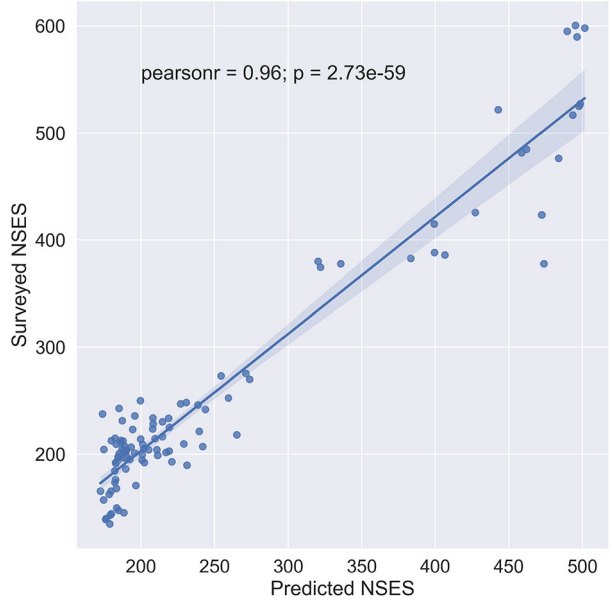

Figure 2 Pearson correlation between predicted and surveyed NSES

ing spectral convolutional filters, to accurately predict the NSES surveyed from the basket graphs of urban neighborhoods in Guayaquil during 2018.

The spatial distribution of the predicted and surveyed NSES in Guayaquil during 2018 is shown in Fig. 3. The maps categorize urban neighborhoods by socioeconomic levels and highlight the geographic trends observed during the quarters of 2018. The urban neighborhoods with the highest NSES are predominantly located in the center and northeast, while the urban neighborhoods with the lowest NSES are found in the south and northwest.

5.2 GNN model explanation

Although explaining the predictions made by a GNN model remains an open problem, various studies have proposed methodologies to address this [26–29]. We use a GNN-Explainer model [27], which takes our trained GNN model and its predictions, returning an explanation in the form of a small subgraph structure of the input graph, along with edge masks that act as attribution explanations. A higher mask value indicates that the corresponding component is more influential in the prediction. We present the ten most influential edges (item combinations) of the subgraph structure returned by the GNN-Explainer for the predictions of the lowest and highest NSES urban neighborhoods in all quarters of 2018 (details are presented in Table 9).

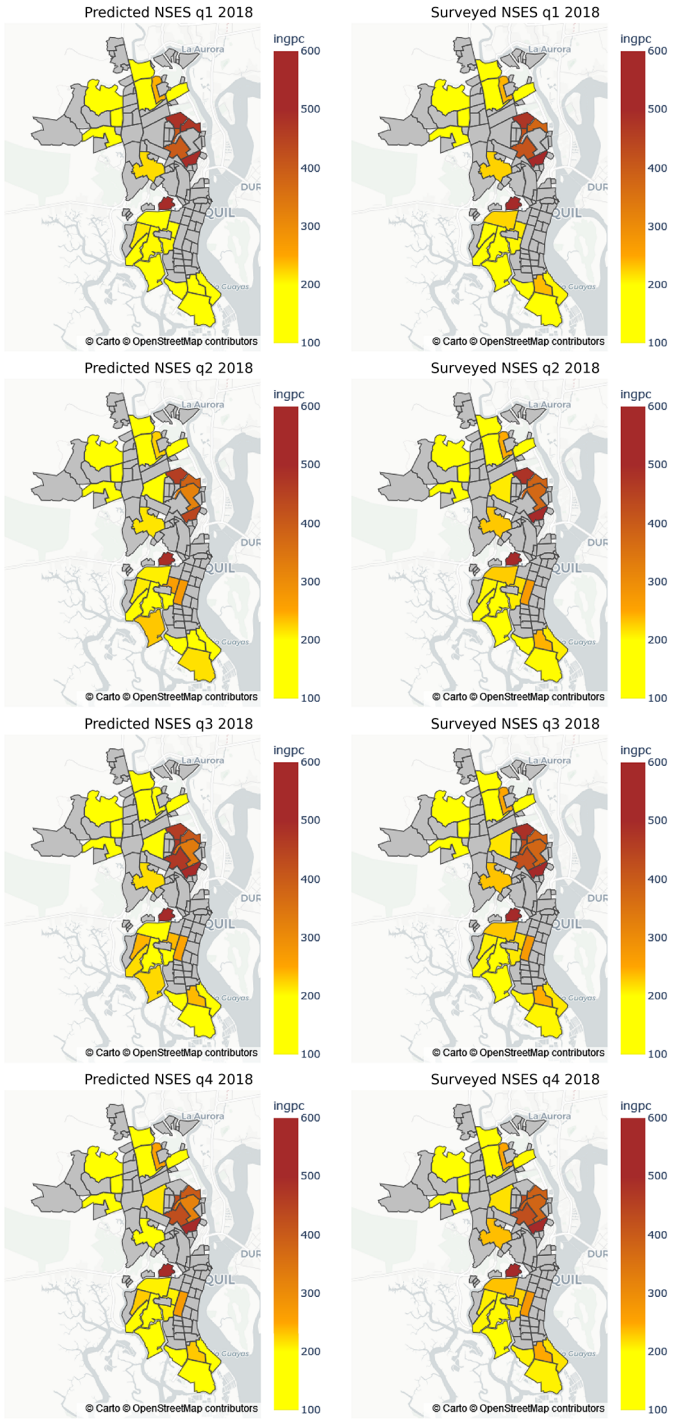

Figure 3 Predicted and surveyed NSES spatial distribution

We analyze the differences in consumption patterns between urban neighborhoods with the highest and lowest NSES. In the urban neighborhood with the lowest NSES, the most influential item combinations in the prediction include disposable tableware and inexpensive yogurt in various quarters. These items reflect the economic reality of the neighborhood, where disposable tableware is commonly used for parties or fast-food businesses,

Table 9 The most common item combinations from the subgraph structure identified by the GNN-Explainer model

Low NSES				High NSES			
$Item_x$	$Price_x$	$Item_y$	$Price_y$	$Item_x$	$Price_x$	$Item_y$	$Price_y$
Quarter 1							
Toilet paper	0.70	Confectionery	0.34	Washing clothes	0.14	Powdered milk	0.62
Confectionery	0.34	***Yogurts***	0.18	Beef	0.54	***Cereals***	0.59
Desserts	0.65	Toilet paper	0.22	Whisky	0.19	Toilet paper	0.13
Flours	0.19	Biscuits	0.39	Bread	0.12	Juices and Tea	0.18
Desserts	0.56	Milk cream	0.10	Washing clothes	0.19	Fruits	0.34
Desserts	0.56	Spreadables	0.24	***Marshmallows***	0.15	Leaf vegetable	0.16
Flours	0.20	Hams	0.18	Juices and Tea	0.10	Chocolates	0.27
Juices and Tea	0.10	Toilet paper	0.22	Washing clothes	0.20	Biscuits	0.29
Liquid cleaners	0.35	Toilet paper	0.22	***Snacks***	0.90	Disp. tableware	0.76
Mortadella	0.15	Washing clothes	0.41	Whisky	0.19	Toilet paper	0.24
Quarter 2							
Sweeteners	0.30	Juices and Tea	0.20	Desserts	0.66	Fruits	0.37
Soft drinks	0.12	Fruits	0.21	Chicken	0.44	Condiments	0.21
Baby perfumes	0.42	Baby perfumes	0.31	Flavored drinks	0.20	Oatmeal	0.19
Hams	0.15	Oils	0.23	Chewing gum	0.43	Yogurts	0.14
Baby perfumes	0.45	Hair treatment	0.79	Liquid cleaners	0.22	Ice cream	0.22
Disp. tableware	0.60	Yogurts	0.50	Chewing gum	0.47	Hair dye	0.23
Beef	0.10	Ladyfingers	0.26	Desserts	0.66	Sweeteners	0.47
Shaving	0.27	***Yogurts***	0.15	Washing clothes	0.25	Soaps	0.28
Bathroom cleaners	0.23	Noodles	0.18	Hams	0.15	Condiments	0.13
Shampoo	0.82	Rice	0.14	Washing clothes	0.41	Fruits	0.32
Quarter 3							
Confectionery	0.39	Toilet paper	0.31	Confectionery	0.28	Rice	0.35
Disp. tableware	0.76	Toilet paper	0.31	Grains	0.10	Condiments	0.17
Pesticides	0.49	Carton milk	0.17	Mortadella	0.20	Desserts	0.56
Washing clothes	0.28	Yogurts	0.54	Mortadella	0.19	Biscuits	0.57
Flavored drinks	0.14	Whisky	0.23	Stains removal	0.37	Kitchen disp.	0.27
Desserts	0.56	Hair dye	0.23	Vodka	0.26	Rice	0.23
Chocolates	0.20	Flours	0.16	Noodles	0.45	Oils	0.16
Liquid cleaners	0.23	Hair dye	0.23	Toilet paper	0.70	Rice	0.15
Foot powder	0.42	Toilet paper	0.31	Mortadella	0.19	***Dog food***	0.10
Chocolates	0.20	***Yogurts***	0.10	Dairy drinks	0.21	Juices and Tea	0.27
Quarter 4							
Whisky	0.17	Toothbrushes	0.21	Washing clothes	0.19	Air fresheners	0.14
Liquid cleaners	0.55	Fruits	0.42	Dried fruits	0.15	Soaps	0.19
Hams	0.10	Hair conditioner	0.58	Sanitary pads	0.29	***Marshmallows***	0.19
Washing clothes	0.28	***Disp. tableware***	0.60	***Marshmallows***	0.19	Coffee	0.39
Body deodorant	0.47	Margarines	0.53	Hams	0.19	***Cereals***	0.28
Candy	0.15	Chocolates	0.27	Baby perfumes	0.37	***Dog food***	0.23
Noodles	0.49	Juices and Tea	0.11	Dairy drinks	0.21	Rice	0.11
Toilet paper	0.25	Toilet paper	0.19	***Snacks***	0.39	Canned foods	0.19
Sweeteners	0.47	Condiments	0.16	Dishwasher det.	0.15	***Dog food***	0.23
Mortadella	0.15	Chocolates	0.20	Dried fruits	0.44	Rice	0.38

and inexpensive yogurts are often consumed for breakfast. In contrast, in the urban neighborhood with the highest NSES, influential item combinations include cereals and marshmallows for breakfast across various quarters. In addition, snacks appear frequently and are generally consumed between meals as part of diet programs. Finally, dog food stands out in several quarters, reflecting the tendency of residents in this neighborhood to purchase processed pet foods.

The results of the GNN-Explainer model confirm that certain items within the item combinations differentiate the consumption behaviors of urban neighborhoods and are linked to NSES.

6 Discussion

Our work focuses on predicting NSES through a data-driven process that includes creating a basket graph from supermarket transactions to represent customers' consumption behavior in an urban neighborhood, followed by training a GNN model to learn the relationship between the basket graph and NSES. The model was evaluated using spectral and spatial convolutional filters with a five-fold cross-validation. The best performance was achieved with spectral convolutional filters, resulting in an $R^2 = 0.91$. The integration of Chebyshev spectral convolutional filters further enhanced this predictive accuracy, suggesting that spectral methods might be particularly effective in capturing structural nuances within neighborhood-level consumption data. This finding underscores the utility of spectral filters in deep learning models aimed at predicting socioeconomic metrics and can guide future work in choosing optimal neural network architectures for similar tasks.

The interpretability of GNNs remains a significant challenge in deep learning. We employed the GNN-Explainer model to identify the most influential item combinations for the NSES predictions, revealing insightful distinctions between high-NSES and low-NSES urban neighborhoods. In urban neighborhoods with lower NSES, items such as disposable tableware and low-cost yogurt were prominent, reflecting economic conditions and certain aspects of lifestyle. In contrast, urban neighborhoods with higher NSES showed a different consumption profile, with influential items such as cereals, marshmallows, and dog food, indicating different dietary habits and consumer preferences. These insights illustrate the socioeconomic diversity of urban neighborhoods, highlighting how digital transaction data can capture meaningful socioeconomic signals.

This work presents a data-driven approach to socioeconomic measurement, which is particularly useful for governments in contexts where traditional data collection is costly and infrequent. By allowing for more frequent updates to socioeconomic status assessments, our method supports informed decisions to design effective resource allocation programs.

7 Limitations and future work

There are some limitations regarding the use of supermarket transactions to predict NSES. First, supermarket transaction data is not publicly available for estimating socioeconomic status, and available public data is often aggregated for privacy reasons. Nevertheless, policymakers could negotiate agreements with supermarket chains to access anonymized transaction data, which could be used with our GNN model to measure NSES with high precision and support decisions for effective resource allocation programs. In addition to access, another limitation is the geographical distribution of supermarkets in the city. Supermarket locations only cover certain urban neighborhoods, making it unfeasible to measure customer consumption behavior through the basket graph in neighborhoods without supermarket stores. However, governments could extend data access agreements to include smaller supermarkets in areas where major chains do not operate.

A further limitation of our methodology is its applicability in rural areas where supermarkets are generally absent. In these cases, governments could rely on census data and

surveys to assess socioeconomic conditions in rural populations. In addition, there are regional limitations in our study, as it was deployed within the urban neighborhoods of a single city. Given that consumption patterns might significantly vary across regions, it would be essential to assess the predictive power of the model using transaction data from other supermarket chains with a different regional presence.

Regarding the NSES data surveyed, there is a limitation due to urban neighborhoods that were not covered by the INEC surveys during the observation period. Consequently, NSES data is unavailable for these neighborhoods, preventing their inclusion in the supervised learning process. Although the GNN model can infer NSES for neighborhoods with supermarkets, we cannot validate these predictions for neighborhoods without surveyed data.

Despite these limitations, our model is adaptable to other countries with access to the data on granular supermarket transactions presented in our study, with the potential to incorporate different poverty proxies that capture human behavior [9, 10, 14] and living conditions [6, 11–13, 30]. The proposed methodology could be implemented as a Software as a Service (SaaS) platform, providing a user-friendly interface that allows users to input data and interpret results without advanced technical skills. This accessibility ensures that the model is available to a wide range of policymakers, enhancing its practical utility for day-to-day decision-making aimed at understanding wealth distribution at a finer spatial granularity than the city level.

Finally, it is essential to note that further studies are needed to improve the NSES estimation. Future research could focus on evaluating the performance of multimodal models in predicting poverty at the intra-urban level by combining the proposed GNN model with various machine learning and deep learning architectures. Using heterogeneous digital data sources could improve NSES predictions, achieving greater predictive power than demonstrated in this study.

8 Conclusions

Our work proposes a data-driven process for predicting NSES using supermarket transactions. We characterize customers' consumption behavior by creating a basket graph that highlights the most common item combinations purchased within an urban neighborhood. Subsequently, we train a GNN model to learn the relationship between basket graph and NSES, evaluating the model's predictive power using spectral and spatial convolutional filers with cross-validation. The best performance was achieved with the Chebyshev spectral convolutional filter, which reached the highest predictive power, with $R^2 = 0.91$ and the lowest $RMSE = 33.70$.

In conclusion, characterizing customer consumption behavior from a graph perspective enables the identification of socioeconomic status at the urban neighborhood level. Furthermore, our GNN model accurately learns the relationship between consumption behavior and NSES. Using spectral convolutional filters, our GNN model offers a new approach to estimating socioeconomic status at the urban neighborhood level, enabling governments to make informed decisions on resource allocation based on the needs of different geographical areas.

Abbreviations

UN, United Nations; SDGs, Sustainable Development Goals; SES, Socioeconomic status; NSIs, National statistical institutes; PCI, Per capita income; IMD, Index of multiple deprivations; CDR, Call detail records; WI, Wealth index; DHS, Demographic

and health survey; NPP-VIIRS, National polar-orbiting partnership visible infrared imaging radiometer suite; NTL, Nighttime lights; OSM, OpenStreetMap; IPI, Integrated poverty index; SRTM, Shuttle radar topography mission; DEM, Digital elevation model; FROM-GLC, Finer resolution observation and monitoring of global land cover; MVGNN, Multi-view graph neural network; GDP, Gross domestic product; LSOA, Lower super output area; QGIS, a geographic information system; INEC, Ecuadorian national institute of statistics and censuses; ENEMDU, National survey of employment, unemployment, and underemployment; NSES, Neighborhood socioeconomic status; GNN, Graph neural network; AGG, Aggregation function; RMSE, Root mean square error; SaaS, Software as a service.

Acknowledgements
We gratefully thank the Escuela Superior Politecnica del Litoral and the Secretaria de Educacion Superior, Ciencia, Tecnologia e Innovacion for financial support.

Author contributions
Study idea: EC, MV, CV. Data analyses: EC. Writing: EC, MV, CV, LE-N, and NV. All authors read and approved the final manuscript.

Funding
The study was funded by the Escuela Superior Politecnica del Litoral and the Secretaria de Educacion Superior, Ciencia, Tecnologia e Innovacion, under the registered research project FIEC-70-02-2018.

Data Availability
The datasets and the code used in this study are available in the "GNN Model to predict NSES (code and data)" repository, https://doi.org/10.6084/m9.figshare.26891407. The poverty map is available on https://vis.csh.ac.at/poverty-maps/.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Escuela Superior Politecnica del Litoral, Guayaquil, Ecuador. [2]Central European University, Vienna, Austria. [3]Complexity Science Hub, Vienna, Austria. [4]Cardiff University, Wales, United Kingdom.

Received: 27 May 2024 Accepted: 5 December 2024 Published online: 17 January 2025

References

1. United Nations, D.o.E., Development, S.A.-S.: (2015). Transforming our world: the 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda
2. Carlsen L, Bruggemann R (2022) The 17 United Nations' sustainable development goals: a status by 2020. Int J Sustain Dev World Ecol 29(3):219–229
3. Lakner C, Mahler DG, Negre M, Prydz EB (2022) How much does reducing inequality matter for global poverty? J Econ Inequal 20(3):559–585
4. Subash SP, Kumar RR, Aditya KS (2018) Satellite data and machine learning tools for predicting poverty in rural India. Agric Econ Res Rev 31(2):231–240
5. Muñetón-Santa G, Manrique-Ruiz LC (2023) Predicting multidimensional poverty with machine learning algorithms: an open data source approach using spatial data. Soc Sci 12(5):296
6. Cao R, Tu W, Cai J, Zhao T, Xiao J, Cao J, Gao Q, Su H (2022) Machine learning-based economic development mapping from multi-source open geospatial data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4:259–266
7. Metzger N, Vargas-Muñoz JE, Daudt RC, Kellenberger B, Whelan TT-T, Ofli F, Imran M, Schindler K, Tuia D (2022) Fine-grained population mapping from coarse census counts and open geodata. Sci Rep 12(1):20085
8. Wardrop N, Jochem W, Bird T, Chamberlain H, Clarke D, Kerr D, Bengtsson L, Juran S, Seaman V, Tatem A (2018) Spatially disaggregated population estimates in the absence of national population and housing census data. Proc Natl Acad Sci 115(14):3529–3537
9. Hernandez M, Hong L, Frias-Martinez V, Whitby A, Frias-Martinez E (2017) Estimating poverty using cell phone data: evidence from Guatemala. World Bank Policy Research Working Paper (7969)
10. Steele JE, Pezzulo C, Albert M, Brooks CJ, Erbach-Schoenberg E, O'Connor SB, Sundsøy PR, Engø-Monsen K, Nilsen K, Graupe B, et al (2021) Mobility and phone call behavior explain patterns in poverty at high-resolution across multiple settings. Humanit Soc Sci Commun 8(1):1–12
11. Zhao X, Yu B, Liu Y, Chen Z, Li Q, Wang C, Wu J (2019) Estimation of poverty using random forest regression with multi-source data: a case study in Bangladesh. Remote Sens 11(4):375
12. Xu Y, Mo Y, Zhu S (2021) Poverty mapping in the dian-gui-qian contiguous extremely poor area of southwest China based on multi-source geospatial data. Sustainability 13(16):8717
13. Li G, Cai Z, Qian Y, Chen F (2021) Identifying urban poverty using high-resolution satellite imagery and machine learning approaches: implications for housing inequality. Land 10(6):648

14. Piaggesi S, Giurgola S, Karsai M, Mejova Y, Panisson A, Tizzoni M (2022) Mapping urban socioeconomic inequalities in developing countries through Facebook advertising data. Front Big Data 5:1006352
15. Fatehkia M, Tingzon I, Orden A, Sy S, Sekara V, Garcia-Herranz M, Weber I (2020) Mapping socioeconomic indicators using social media advertising data. EPJ Data Sci 9(1):22
16. Vinkeles Melchers NV, Gomez M, Colagiuri R (2009) Do socio-economic factors influence supermarket content and shoppers' purchases? Health Promotion Journal of Australia 20(3):241–246
17. Drewnowski A, Moudon AV, Jiao J, Aggarwal A, Charreire H, Chaix B (2014) Food shopping behaviors and socioeconomic status influence obesity rates in Seattle and in Paris. Int J Obes 38(2):306
18. Pechey R, Monsivais P (2015) Supermarket choice, shopping behavior, socioeconomic status, and food purchases. Am J Prev Med 49(6):868–877
19. Bannister A, Botta F (2021) Rapid indicators of deprivation using grocery shopping data. R Soc Open Sci 8(12):211069
20. Uruchima J, Renehan C, Castro N, Cevallos W, Levy K, Eisenberg JN, Lee GO (2023) A qualitative study of food choice in urban coastal Esmeraldas Ecuador. Current Developments in Nutrition 7(5):100093
21. Comercio E (2023) Tiendas compiten con más supermercados en barrios. https://www.elcomercio.com/actualidad/negocios/tiendas-compiten-con-supermercados-barrios.html. Accessed: 2024-02-10
22. Rivadeneira D, et al (2018) Encuesta nacional de empleo, desempleo y subempleo (enemdu) documento metodologico. Instituto Nacional de Estadística y Censos INEC. Recuperado de: https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Septiembre-2018/ENEMDU_Metodologia%20Encuesta%20Nacional%20de%20Empleo%20Desempleo%20y%20Subempleo.pdf
23. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp 1188–1196. PMLR
24. Videla-Cavieres IF, Ríos SA (2014) Extending market basket analysis with graph mining techniques: a real case. Expert Syst Appl 41(4):1928–1936
25. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in neural information processing systems, vol 29
26. Luo D, Cheng W, Xu D, Yu W, Zong B, Chen H, Zhang X (2020) Parameterized explainer for graph neural network. Adv Neural Inf Process Syst 33:19620–19631
27. Ying Z, Bourgeois D, You J, Zitnik M, Leskovec J (2019) Gnnexplainer: Generating explanations for graph neural networks. In: Advances in neural information processing systems, vol. 32
28. Schlichtkrull MS, De Cao N, Titov I (2020) Interpreting graph neural networks for nlp with differentiable edge masking. Preprint. arXiv:2010.00577
29. Zhang J, Chen Z, Mei H, Luo D, Wei H (2023) Regexplainer: Generating explanations for graph neural networks in regression task. Preprint. arXiv:2307.07840
30. Ma J, Yang L, Feng Q, Zhang W, Yu PS (2023) Graph-based village level poverty identification. In: Proceedings of the ACM web conference 2023, pp 4115–4119

Publisher's Note