

# The Relation of Crowdsourced Data, Traffic Congestion Patterns, and Road Network in Samarinda City, Indonesia

**Arief Hidayat**

Urban and Regional Planning, Institut Teknologi Kalimantan, Indonesia  
arief.hidayat@lecturer.itk.ac.id (corresponding author)

**Muhammad Fadhillah Nabhan Ednolian**

Urban and Regional Planning, Institut Teknologi Kalimantan, Indonesia  
fadhilnabhan768@gmail.com

**Mohtana Kharisma Kadri**

Urban and Regional Planning, Institut Teknologi Kalimantan, Indonesia  
mohtana.kharisma@lecturer.itk.ac.id

**Happy Aprilia**

Electrical Engineering, Institut Teknologi Kalimantan, Indonesia  
happy.apriliah@lecturer.itk.ac.id

**Umar Mustofa**

Urban and Regional Planning, Institut Teknologi Kalimantan, Indonesia  
Umar.mustofa@lecturer.itk.ac.id

**Bimo Aji Widyantoro**

Urban and Regional Planning, Institut Teknologi Kalimantan, Indonesia  
bimo.widyantoro@lecturer.itk.ac.id

**Hijriah**

Civil Engineering, Institut Teknologi Kalimantan, Indonesia  
hijriah@lecturer.itk.ac.id

*Received: 30 September 2025 | Revised: 8 November 2025, 16 November 2025, and 20 November 2025 | Accepted: 21 November 2025*

*Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.15232>*

## ABSTRACT

This study analyzes traffic congestion in Samarinda City using a crowdsourcing approach based on real-time Google Maps data to identify congestion patterns and contributing factors. The research examines the relationship between road network characteristics, including road hierarchy, road width, intersections, and traffic density, across different urban road segments using a descriptive and quantitative approach. Secondary data were collected via the Google Maps Application Programming Interface (API) and processed in ArcGIS to produce congestion hotspot maps using the Getis-Ord  $G_i^*$  spatial analysis method. The results indicate that between 11 and 17 August 2025, the highest congestion occurred on the Samarinda-Bontang Road. Intersections were identified as the most significant variable influencing congestion levels. Based on these findings, recommendations were made to improve the road network and manage intersections to reduce congestion in Samarinda City. This study also to urban planning by demonstrating the potential of crowdsourced data as an efficient tool for data collection and analysis to support sustainable transportation policies.

*Keywords-traffic congestion; crowdsourcing; spatial analysis; road network*

## I. INTRODUCTION

Samarinda City is the capital of East Kalimantan Province and has the largest population in Kalimantan, with 861,878 inhabitants and an area of 71,800 ha, according to 2022 data from the Samarinda City Government. As the provincial capital, Samarinda supports various activities throughout East Kalimantan. It serves not only as an administrative center but also as an economic hub that supports provincial-scale activities. This condition indicates that Samarinda functions as both a destination and a transit hub, resulting in relatively high, steadily increasing community mobility. However, transportation management in Samarinda has not yet met residents' expectations for daily mobility. According to the Samarinda City Medium-Term Development Plan, one of the key transportation issues is traffic congestion.

The Transportation Masterplan also notes that congestion occurs widely in several parts of Samarinda. According to Law No. 22 of 2009 on Roads, traffic, and transportation, congestion is defined as a reduction in the smoothness of traffic flow, generally caused by insufficient road capacity relative to the number of vehicles, roadside obstacles, or accidents. Congestion is influenced by several factors, such as vehicle size, driving behavior, pedestrian activity, weather conditions, and road network issues, including inadequate capacity and numerous intersections [1]. According to the Samarinda City Transportation Agency's 2025 Medium-Term Development Plan, there are 20 congestion points across various road segments in the city.

The Samarinda City Transportation Masterplan highlights several causes of congestion, including limited road networks and the lack of connectivity between roads according to their functional hierarchy or road class. The city's road network spans 898.024 km. Congestion occurs on a road segment when the degree of saturation exceeds 0.5 [2]. The master plan states that several road segments experience saturation above 0.5, which is categorized as congested. These include Jl. Dr. Soetomo-Jl. Pahlawan, Jl. Slamet Riyadi, Jl. Jembatan Mahakam, Jl. RE Martadinata, Jl. Meranti, Jl. Tengawang, Jl. Bung Tomo, Jl. Yos Sudarso-Jl. Gajah Mada, Jl. Awang Long B, Jl. Ir. H. Juanda, Jl. A. Wahab Syahrani, and several other roads.

Transportation data collection can be very costly and time-consuming. Internet-based data have become a powerful and more efficient alternative for capturing information. With advances in information and communication technology, data can now be easily accessed through internet-connected applications. The Internet of Things (IoT) refers to a network of interconnected physical devices equipped with sensors. This includes a wide range of devices, such as mobile phones, wearable technologies, industrial equipment, and various electronic systems [3]. One technology that can identify congestion points in an area is Google Maps. Google Maps is a web-based mapping and navigation service developed by Google that offers real-time traffic information, including road conditions, congestion, and estimated travel times. This technology enables users to plan their journeys more

effectively and avoid congested routes. The Google Traffic feature in Google Maps provides a quick overview of congestion levels in a city [4]. Google Maps is often used to identify congestion-prone points during specific hours at key activity locations and the results are used to formulate transportation policy recommendations.

In the current study, Google Maps is integrated with a crowdsourcing approach. Crowdsourcing is a mechanism for analyzing and extracting common elements to identify the fundamental characteristics of a task [5]. It typically involves the participation of a large group or organization through the internet to achieve outcomes that cannot be obtained by an individual or a small group [6]. Crowdsourcing can, therefore, be defined as a concept or system derived from contributions made by a large number of participants. Previous studies have demonstrated the use of crowdsourcing to collect application-based data from community reports on real-time road conditions in Malang City. In this research, crowdsourcing is used to identify congestion points in Samarinda using real-time data from Google Maps. The advantages of this approach include speed, quality, flexibility, and diversity.

## II. METHODOLOGY

### A. Research Approach and Type

This study adopts a quantitative, positivist approach. It examines a specific population or sample through random sampling, collects data using research instruments, and analyzes them statistically [7]. The research employs a descriptive approach as its primary type. Descriptive research aims to examine an event or phenomenon by asking individuals or groups to describe their experiences or observations of that phenomenon [7]. This study also utilizes Geographic Information Systems (GIS), a method commonly used for spatial research. GIS is a computer-based information system for managing spatially referenced data. It collects, stores, manages, analyzes, displays, and applies geographic information, and serves as a general tool for analyzing and managing large-scale spatial data [8]. In this research, GIS was employed to analyze the crowdsourced data to identify traffic congestion patterns.

### B. Analyzing Traffic Congestion

The variables used in the study were designed to analyze congestion points indicated by increases in traffic density through the crowdsourcing method. Table I presents the research variables for Objective 1.

TABLE I. RESEARCH VARIABLES FOR OBJECTIVE 1

Objective	Indicator	Variable
Analyzing congestion points marked by an increase in traffic density using crowdsourcing	Congestion points	Distribution of location points
	Data collection through crowdsourcing	Total number of congestion incidents collected from Google Maps

1) Data Collection Method

The data collection method refers to the technique or approach used to obtain valid information relevant to the research problem. In this study, only secondary data obtained from institutions or organizations were used. The purpose of conducting a secondary survey was to analyze congestion points identified by increased traffic density using crowdsourcing. This method was applied to collect congestion data in Samarinda City from Google Maps [10]. Table II presents data collection techniques used for Objective 1.

TABLE II. DATA COLLECTION METHOD FOR OBJECTIVE 1

Data type	Data source	Data collection technique
Distribution of congestion points	Google Maps	Secondary survey

The data collection stages using the crowdsourcing method are portrayed in Figure 1.

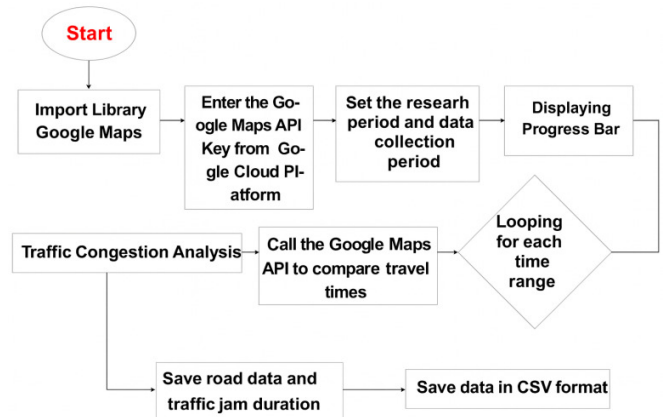


Fig. 1. Data collection stages for Objective 1.

Tables III and IV describe the data collection process based on several general steps.

The CSV data resulting from processing based on the flowchart were entered into ArGIS or the GIS program, which processed the data for further analysis.

TABLE III. FIRST GENERAL STEP FOR GOOGLE API

Steps	Process	Information
1	API code retrieval process	Open the Google Arcade web
2		After logging in, sign in using personal email
3		Enter the access code that is available on the Google Arcade
4		Select one of the modules and click Start Module
5		After the display appears as in the image, start the lab
6		Log in to the Google Cloud. The console uses the designated email, which remains active for the specified period
7		After logging in, the display is shown in the image, then click the three lines in the top left corner
8		Select APIs and Services, then select the library.
9		Create a project.
10		In the Library search bar, search for Routes API
11		After selecting, click the enable button
12		Select the project that was created earlier
13		Once the required API code appears, click copy
14		Paste the key into the script under API configuration

TABLE IV. SECOND GENERAL STEP FOR RETRIEVAL OF API ROAD DATA

Script code	Function
import osmnx US ox import pandas US pd	OSMNx is used to download geographic data from OpenStreetMap (OSM) [11], specifically roads, buildings, networks, etc. Pandas is used to manipulate and store data in tabular format (DataFrame).
# Get road data for Samarinda place_name = " Samarinda, East Kalimantan , Indonesia"	Determines the location for network data collection
graph = ox.graph_from_place ( place_name , network_type = ' drive ' )	Collects data from OSM according to the location name and takes road network data that can only be accessed by vehicles
streets = ox.graph_to_gdfs ( graph , nodes = False , edges = True )	Converts a graph into a GeoDataFrame of road sections
data = {" name " : streets [ ' name ' ] . fillna ( " Unnamed Road " ) . tolist ( ) , " geometry " : [ { " coords " : line.coords [ : ] } for line in streets [ ' geometry ' ] ]	Converts GeoDataFrame to format Python dictionary

TABLE V. THIRD GENERAL STEP FOR RETRIEVAL OF API CONGESTION

Script	Function
<pre>import osmnx US ox import pandas US pd import requests import csv import OS from datetime import datetime, time delta from TQDM import TQDM from shapely.geometry import LineString import random</pre>	<p>osmnx: fetches road data from OSM  pandas: processes data in tabular form  requests: to call the Google Routes API  csv: used to create a CSV file  os: checks whether the file exists  datetime, timedelta: creates API departure time  tqdm: creates a progress bar while the process is running  shapely: processes LineString geometry  random: used to create jitter (small modifications to coordinates)</p>
<pre>API_KEY = "... " # Replace with your API Key place_name = "Samarinda, East Indonesia, Indonesia" sample_fraction = 1 output_csv = 'SMD_traffic_segmented.csv' congestion_csv = 'Samarinda_congestion_points.csv' ROUTES_API_URL = "https://routes.googleapis.com/directions/v2:computeRoutes" time_range = [ "08:00 - 09:00" ] # Time range for 08:00 and 12:00 now = datetime.now ()</pre>	<p>API_KEY: specifies Google API key  place_name: determines the city to be analyzed  time_range: determines the analysis time  output_csv: specifies the output name CSV file  sample_fraction =1: uses 100% of the road data  now: used as the date for the log</p>
<pre>graph = ox.graph_from_place ( place_name , network_type = ' drive ' ) streets = ox.graph_to_gdfs ( graph , nodes = False , edges = True )</pre>	<p>Takes the Samarinda City Road Network  Converts road data to GeoDataFrame (edges/roads only)</p>
<pre>big_road_types = [ " motorway " , " trunk " , " primary " , " secondary " , " tertiary " ] def is_big_road ( hwy ) : if isinstance ( hwy , list ) : return any ( h in big_road_types for h in highway ) return highway in big_road_types streets = streets [ streets [ " highway " ] . apply ( is_big_road ) ] df_streets = streets [ [ ' name ' , ' geometry ' ] ] . copy ( ) df_streets [ ' name ' ] = df_streets [ ' name ' ] . fillna ( " Unnamed Road " ) df_streets = df_streets [ df_streets.geometry.type == ' LineString ' ] df_streets = df_streets.sample ( frac = sample_fraction , random_state = 42 ) . reset_index ( drop= True )</pre>	<p>Filters only major road networks and prepares road data</p>
<pre>if not os.path.exists ( output_csv ): with open ( output_csv , 'w' , newline = '' ) as f : writer = csv.writer ( f ) writer.writerow ( [ "Date" , "Time (Range)" , "Street Name" , "Initial Coordinates" , "Final Coordinates" , "Distance (m)" , "Duration ( min )" , " Delay per 100m ( min )" , "Congestion Level" ] )</pre>	<p>Creates a Master CSV file with headers date, time, street name, start and end coordinates, distance, duration, delay, and congestion level</p>
<pre>if not os.path.exists ( congestion_csv ): with open ( congestion_csv , 'w' , newline = '' ) as f : writer = csv.writer ( f ) writer.writerow ( [ "Date" , "Time (Range)" , "Street Name" , "Initial Congestion Coordinates" , "Final Congestion Coordinates" , "Step Distance (m)" , "Step Duration ( min )" , " Delay per 100m ( min )" ] )</pre>	<p>Collects the congestion data</p>

Script	Function
<pre>def split_linestring_by_length ( line , max_length = 200 ) : segments = [] total_length = line.length start_distance = 0 while start_distance &lt; total_length : end_distance = min ( start_distance + max_length , total_length ) p1 = line.interpolate ( start_distance ) p2 = line.interpolate ( end_distance ) segments.append ([ ( p1.x , p1.y ) , ( p2.x , p2.y )]) start_distance = end_distance return segments</pre>	<p>Divides the road network per 200 m</p>
<pre>def jitter_coords ( coords , jitter_meters = 3 ) : jitter_degree = jitter_meters / 111320 # convert meters to degrees return ( coord [ 0 ] + random.uniform ( - jitter_degree , jitter_degree ) , coord [ 1 ] + random.uniform ( - jitter_degree , jitter_degree )</pre>	<p>Addes jitter to get the variation of coordinates so that no road network is identical</p>
<pre>def get_route_details ( start_coord , end_coord , departure_time ) : payload = { " origin " : { " location " : { " latLng " : { " latitude " : start_coord [ 1 ] , " longitude " : start_coord [ 0 ]}}}, " destination " : { " location " : { " latLng " : { " latitude " : end_coord [ 1 ] , " longitude " : end_coord [ 0 ]}}}, " travelMode " : "DRIVE" , " routingPreference " : "TRAFFIC_AWARE" , " departureTime " : departure_time.isoformat () + "Z" , " extraComputations " : [ "TRAFFIC_ON_POLYLINE" ] } headers = { " Content-Type " : " application / json " , "X- Goog -Api- Key " : API_KEY , "X- Goog - FieldMask " : " routes.duration,routes.distanceMeters,routes.legs.steps " } try : response = requests.post ( ROUTES_API_URL , json = payload , headers = headers ) response.raise_for_status () data = response.json () if " routes " in data and len ( data [ " routes " ]) &gt; 0 : routes = data [ " routes " ][ 0 ] total_duration = int ( route.get ( " duration " , "0s" )[: -1 ]) / 60 total_distance = route.get ( " distanceMeters " , 0 ) steps = route.get ( " legs " , [{}])[ 0 ] . get ( " steps " , []) return total_duration , total_distance , steps except Exception as e : print ( f "API error : { e } " ) return None , None , []</pre>	<p>Sends a request to the Google API to request travel duration, distance, and congestion points</p>

Script	Function
<pre> With tqdm ( total= len ( df_streets ) * len ( time_range ), desc = " Processing Segments " ) US pbar : for _ , row in df_streets.iterrows ( ): street_name = row [ ' name ' ] geometry = rows [ ' ' geometry ] segments = split_linestring_by_length ( geometry )  for t_idx , time_slot in enumerate ( time_range ): hour = int ( time_slot.split ( ":" ) [ 0 ] ) departure_time = now . replace ( hour = hour , minute = 0 , second = 0 , microsecond = 0 ) + timedelta ( days = 1 )  for start_coord , end_coord in segments : # Apply jitter to differentiate each time start_jittered = jitter_coords ( start_coord , jitter_meters = ( t_idx + 1 ) * 2 ) end_jittered = jitter_coords ( end_coord , jitter_meters = ( t_idx + 1 ) * 2 )  total_duration , total_distance , steps = get_route_details ( start_jittered , end_jittered , departure_time )  if total_duration is not None and total_distance &gt; 0 : delay_per_100m = ( total_duration / total_distance ) * 100 congestion = "Low" if delay_per_100m &gt; 2.5 : congestion = "Traffic jam" elif delay_per_100m &gt; 1 : congestion = "Medium"                     </pre>	<p>The main process for determining the congestion point is that if the delay obtained is greater than 2.5 minutes, then it is said to be congested, if the delay is greater than 1 but does not exceed 2.5 min, it is said to be medium, and if the delay is under 1 min, then it is said to be smooth</p> <p>Loops every path and every process Loops every time Determines departure time</p>
<pre> with open ( congestion_csv , 'a' , newline = ' ' ) as f : writer = csv.writer ( f ) writer.writerow ( [ now.strftime ( "%Y-%m-%d " ) , time_slot , street_name , step_start , step_end , round ( step_distance , 1 ) , round ( step_duration , 2 ) , round ( step_delay_per_100m , 2 )                     </pre>	<p>Saves results to the main CSV file</p>
	<p>CSV results</p>

2) Data Validation

Data validation is the process of comparing actual field conditions with congestion data collected through

crowdsourcing on Google Maps. The validation or verification process was conducted using primary surveys at congestion distribution points across the city to ensure that Google Maps data matched real conditions. Data validation was carried out

during peak hours, specifically from 07:00 to 08:00, 12:00 to 13:00, and 17:00 to 18:00. These time periods were chosen because road users are typically engaged in activities such as commuting to work, attending school, or travelling for other daily purposes during these hours.

TABLE VI. FOURTH GENERAL STEP FOR GIS ANALYSIS PROCESS

Steps	Information
1	Open the GIS Application
2	Select the file, then select Add data and select XY data
3	Then import the CSV that already has coordinates, and find out the X and Y coordinates
4	After the point appears, export the data to save as a shapefile.
5	Search hotspot analysis
6	In the Input Feature, select the point shapefile. In the Input Field, choose the delay field. For Output, choose a folder for storing the results. Under Conceptualization of Spatial Relationship, select Fixed Distance Band, and for Distance Method, choose Euclidean Distance
7	Finished hotspot map

### C. Relation between Traffic Congestion and Road Network

#### 1) Descriptive Analysis of Congestion Characteristics

Descriptive analysis is an analytical technique used to describe or present collected data based on actual conditions, without making generalizations or predictions [7]. Generally, descriptive analysis focuses on real data and explains phenomena as they occur in the field. For Objective 1, descriptive analysis was conducted to interpret congestion conditions in Samarinda City using congestion point data collected from Google Maps.

#### 2) Spatial Analysis

Spatial analysis using ArcGIS software was used to visualize congestion hotspots and congestion intensity in Samarinda City. The Getis-Ord  $G_i^*$  method was applied in this stage. This technique is a spatial statistical method used to identify and evaluate statistically significant spatial clusters or patterns in geographic data. The process measures the extent to which variable values exhibit spatial clustering [9]. Tables VII and VIII present data collection methods deployed for Objective 2, which is to analyze land-use characteristics and their relationship to congestion in Samarinda City. The secondary data were obtained from surveys conducted by relevant government agencies.

#### 3) Regression

Linear regression is a statistical technique used to model and analyze the relationship between a quantitative dependent variable and one or more independent variables. It uses a linear equation to minimize prediction errors. Linear regression aims to identify a linear association between the dependent variable ( $Y$ ) and the independent variable(s) ( $X$ ). In this study, linear regression was employed to determine which variable most significantly influences traffic congestion in Samarinda City. The dependent variable ( $Y$ ) represents delay, while the independent variables ( $X$ ) represent intersections, road width, and road hierarchy.

TABLE VII. DATA COLLECTION FOR OBJECTIVE 2

Data type	Data source	Data collection technique
Analyzing the impact of the road network on traffic congestion in Samarinda City	Road network	Road hierarchy
		Road width
		Intersections

TABLE VIII. DATA COLLECTION METHOD FOR OBJECTIVE 2

Data type	Data source	Data collection technique
Road hierarchy	SHP Samarinda City road decree	Secondary survey
Road width		
Intersections		

## III. RESULTS

### A. Congestion Data

Data collection using the crowdsourcing method was conducted from 11 August 2025 to 17 August 2025, between 07:00 Central Indonesian Time (WITA) and 18:00 WITA. A total of 274,358 traffic data points were collected and categorized into three types: low, moderate, and congested. Within this dataset, 798 congestion points were identified in Samarinda City during the observation period. The collected data included the date of collection (within a 1-h window), the road name at each data location, the distance from the starting point to the endpoint, the normal travel time from Google Maps, the delay due to congestion, and the congestion level.

### B. Road Hierarchy

In Samarinda City, several road segments experience congestion across different road hierarchies, including primary arterial roads, primary collector roads, secondary collector roads, secondary local roads, primary local roads, and secondary local roads. Table IX presents the number of road segments by hierarchy in Samarinda City, as stated in the Samarinda City Mayor's Decree. Based on Table IX, there are 18 segments of Primary Arterial Roads, three segments of Secondary Arterial Roads, 31 segments of Primary Collector Roads, 78 segments of Secondary Collector Roads, 11,206 segments of Secondary Local Roads, two segments of Primary Local Roads, and 470 segments of Secondary Local Roads.

### C. Intersections

Intersections are one of the variables used in this study to examine the effect of intersection presence on traffic congestion in Samarinda City. Intersection data were obtained from the road network shapefile and analyzed spatially using the intersect feature to identify intersection points. Based on the spatial analysis, 30,515 intersection points were identified within Samarinda City. Figure 2 shows the distribution of intersections along road segments in Samarinda City, along with congestion data for both present and absent intersections. The data were recorded between 11 August 2025 and 17 August 2025. Based on data collected from 11 August 2025 to 17 August 2025, a total of 536 congestion points were identified at intersections, while 222 occurred at locations without intersections, as presented in Table X.

TABLE IX. ROAD HIERARCHY

Road hierarchy	Number of segments	Source
Primary arterial road	18	KEPMEN PUPR NO. 430/KPTS/M/2022
Secondary arterial road	3	Samarinda City Road Status Decree
Primary collector road	31	Analysis and Planning Results 2019, Satellite Imagery Interpretation Results (Pleiades and SPOT 7), National Data
Secondary collector road	78	Analysis and Planning Results 2019, Satellite Imagery Interpretation Results (Pleiades and SPOT 7), National Strategic Projects and Proposed Revision of the East Kalimantan Rencana Tata Ruang Wilayah (RTRW or Regional Spatial Plan) 2020-2040 (2020), East Kalimantan RTRW 2016-2036 (2020), Samarinda City Road Status Decree No. 622.3/228/HK-KS/IV/2017
Secondary local road	11,206	Analysis and Planning Results Satellite Imagery Interpretation
Primary local road	2	Satellite Imagery Interpretation No. 622.3/228/HL-KS/IV/2017, Satellite Imagery Interpretation Result (Pleiades & SPOT 7), Analysis and Planning Results 2019
Secondary local road	470	Satellite Imagery Interpretation Results (Pleiades and SPOT 7), Analysis and Planning Results 2019
		Samarinda City Road Status Decree No. 622.3/228/HK-KS/IV/2017, Satellite Imagery Interpretation Results (Pleiades and SPOT 7), Analysis and Planning Results 2019

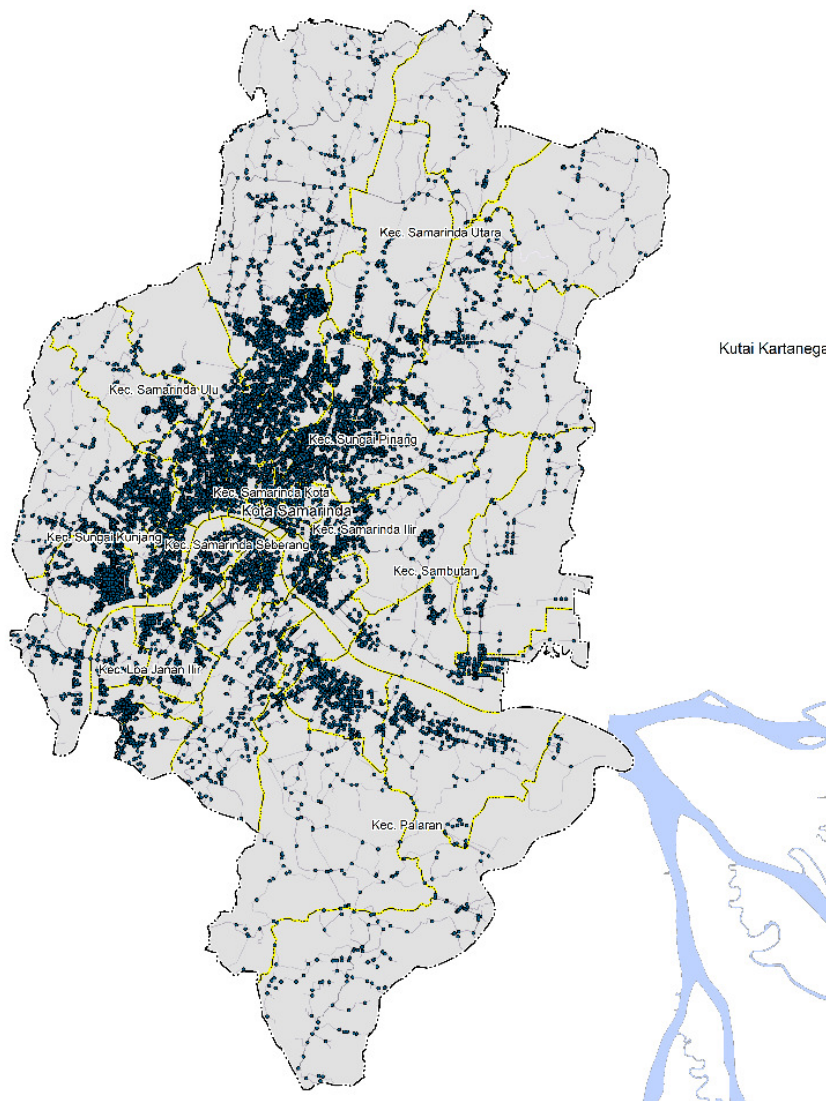


Fig. 2. Intersections in Samarinda city.

TABLE X. NUMBER OF POINTS BASED ON INTERSECTION

Intersection	Number of congestion points
Present	536 points
Not present	222 points

D. Traffic Congestion Locations

From the data collected using the crowdsourcing method, an overlay was performed with road classes defined by the Samarinda city road decree to determine which classes experienced the most congestion. Table XI presents the results of overlaying congestion points with road hierarchy classes.

TABLE XI. NUMBER OF POINTS BASED ON ROAD HIERARCHY

Road hierarchy	Number of congestion points
Primary arterial road	230 points
Secondary arterial road	4 points
Primary collector road	60 points
Secondary collector road	66 points
Secondary neighborhood road	424 points
Secondary local road	86 points

As displayed in Table XI, traffic congestion in Samarinda City during the week of 11-17 August 2025 was dominated by primary collector roads. This was due to the large number of land uses along these roads, which contributed to increased vehicle traffic and congestion.

E. Data Validation

Based on field validation or ground truth, several congestion points were identified and chosen as study samples.

1) Morning Observations (07.00 – 08.00 WITA)

The morning time slot, typically associated with rush hour, was documented on seven different streets across the week. These streets included Jl. KH. Abdul Hasan (Aug 11), Jl. Suryanata (Aug 12), Jl. PM Noor (Aug 13), Jl. MT Haryono (Aug 14), Jl. Lambung Mangkurat (Aug 15), Jl. Marsda A. Saleh (Aug 16), and Jl. Cipto Mangunkusumo (Aug 17). The ground truth shows various conditions, from traffic on wet roads on August 11 to moderate traffic along built-up areas on the other days. The observation on Jl. Lambung Mangkurat on August 15 indicated particularly heavy traffic, possibly due to a constrained roadway. Every morning record was verified as Valid.

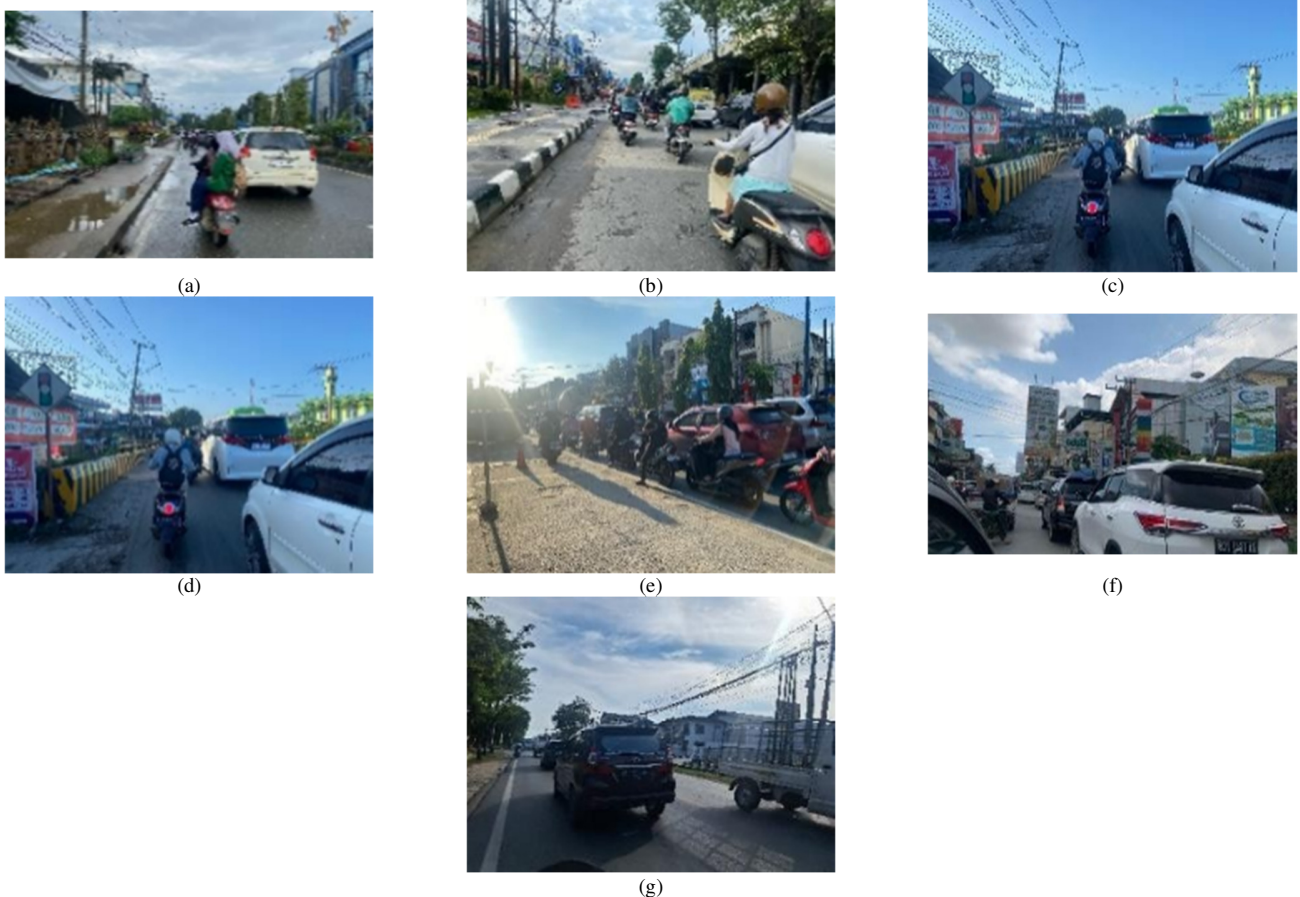


Fig. 3. Midday observations: (a) GT 11 August, 2025; (b) GT 12 August, 2025; (c) GT 13 August, 2025; (d) GT 14 August, 2025; (e) GT 15 August, 2025; (f) GT 16 August, 2025; (g) GT 17 August, 2025.

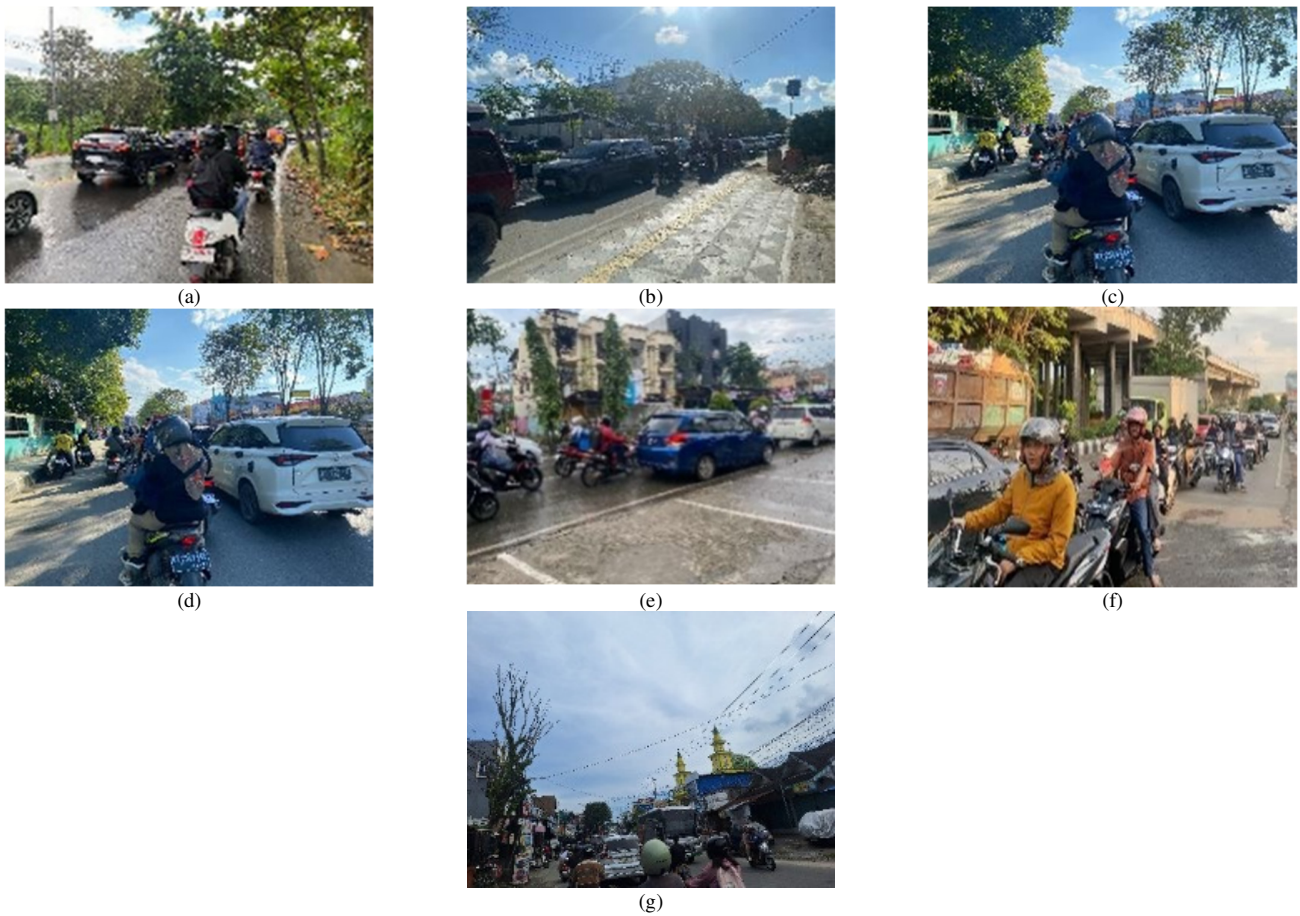


Fig. 4. Late afternoon/evening observations: (a) GT 11 August, 2025; (b) GT 12 August, 2025; (c) GT 13 August, 2025; (d) GT 14 August, 2025; (e) GT 15 August, 2025; (f) GT 16 August, 2025; (g) GT 17 August, 2025.

2) Late Afternoon/Evening Observations (16.00 – 17.00 WITA)

The late afternoon slot, covering the potential evening rush, was consistently monitored across all six weekdays from August 11 to 16, and included streets such as JL. Abdul Wahab Syhramie (Aug 11), Jl. Bontang-Samarinda (Aug 12), Jl. KH Harun Nafsi (Aug 13), Jl. Mayjen Soetoyo (Aug 14), Jl. Agus Salim (Aug 15), and Jl. Samarinda-Bontang (Aug 16). The ground truth generally suggests high traffic volumes, with specific congestion documented on Jl. KH Harun Nafsi and Jl. Mayjen Soetoyo. The documentation for Jl. Samarinda-Bontang showed a significant number of motorbikes near an elevated structure. All collected data for the late afternoon time slot across the six days were validated.

F. Hotspot Analysis

The Hot Spot Analysis (Getis-Ord  $G_i^*$ ) was used to determine the frequency and spatial distribution of traffic congestion in Samarinda City. The congestion points used in this analysis were obtained from the crowdsourced dataset. The study was conducted from 11 August 2025 to 17 August 2025. Congestion frequency was calculated based on the delay-time value per 100 m.

TABLE XII. HOTSPOT ANALYSIS

Description	Number of points
Cold spot – 99% confidence	83,467
Cold spot – 95% confidence	41,894
Cold spot – 90% confidence	25,178
Not significant	192,370
Hotspot – 90% confidence	15,133
Hotspot – 95% confidence	26,361
Hotspot – 99% confidence	139,411

Based on the results of the hotspot analysis conducted from 11 to 17 August 2025, several cold- and hotspot areas related to traffic congestion were identified. A total of 83,467 points were classified as cold spots with 99% significance, 41,894 points as cold spots with 95% significance, and 25,178 points as cold spots with 90% significance. Meanwhile, 192,370 points were found to be statistically not significant. In contrast, hot spots were also detected, comprising 15,133 points at 90% significance, 26,361 points at 95% significance, and 139,411 points at 99% significance. These results illustrate the spatial distribution and intensity of congestion within the city. Figure 3 illustrates the pattern of traffic congestion across Samarinda City.

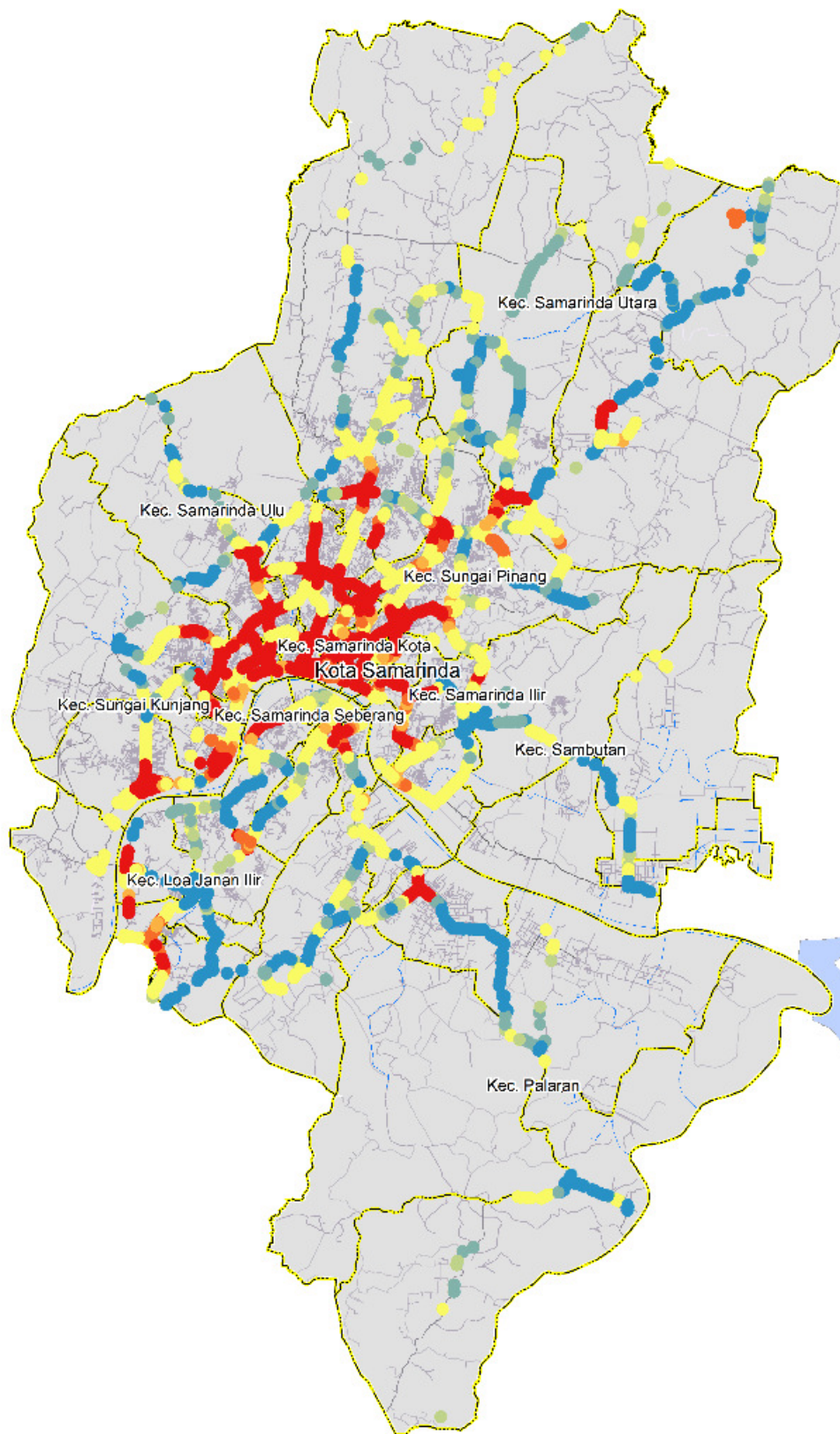


Fig. 5. Hotspot analysis of Samarinda city.

G. Regression Analysis

After collecting congestion data using the crowdsourcing method and performing hotspot analysis of the congestion points, the congestion data were analyzed in the SPSS application using regression analysis. Table XIII portrays the regression analysis results. Based on the regression analysis conducted from 11 to 17 August 2025, the p-values were below 0.05, indicating that the variables had a statistically significant effect on traffic congestion in Samarinda City. On 11 August 2025, intersections and road hierarchy significantly influenced

congestion levels. On 12 August 2025, intersections and road width were found to have a significant effect. On 13 August 2025, intersections and road hierarchy again showed a significant relationship with congestion. On 14 August 2025, all three variables (intersections, road hierarchy, and road width) had a significant influence, and the same pattern was observed on 15 August 2025. On 16 August 2025, intersections and road width remained significant, while on 17 August 2025, intersections and road hierarchy were once more found to have a significant effect on congestion.

TABLE XIII. REGRESSION ANALYSIS

Time	Road network					Equation
	Regression statistical performance					
		Coefficients	Standard error	t-stat	p-value	
11 August	Intercept	0.274483	0.006124	44.8232	0	$Y = 0.274483 + 0.033227.X_1$ $- 0.00204.X_2$ $- 0.00063.X_3$
	Intersection	0.033227	0.002077	16.00115	1.57E-57	
	Road hierarchy	-0.00204	0.000708	-2.88439	0.003923	
	Road width	-0.00063	0.000506	-1.24824	0.211948	
12 August	Intercept	0.272107	0.005846	46.54612	0	$Y = 0.272107 + 0.032464.X_1$ $- 0.00103.X_2$ $- 0.00115.X_3$
	Intersection	0.032464	0.001982	16.37886	3.46E-60	
	Road hierarchy	-0.00103	0.000637	-1.61028	0.107342	
	Road width	-0.00115	0.00049	-2.34098	0.019236	
13 August	Intercept	0.273316	0.006085	44.91771	0	$Y = 0.273316 + 0.033088.X_1$ $- 0.00152.X_2$ $- 0.00077.X_3$
	Intersection	0.033088	0.002062	16.0435	7.95E-58	
	Road hierarchy	-0.00152	0.000662	-2.29553	0.021706	
	Road width	-0.00077	0.00051	-1.51222	0.130482	
14 August	Intercept	0.285479	0.005987	47.67963	0	$Y = 0.285479 + 0.033178.X_1$ $- 0.00259.X_2$ $- 0.00159.X_3$
	Intersection	0.033178	0.002062	16.08677	3.97E-58	
	Road hierarchy	-0.00259	0.000697	-3.71762	0.000201	
	Road width	-0.00159	0.0005	-3.17513	0.001498	
15 August	Intercept	0.288205	0.006133	46.99442	0	$Y = 0.288205 + 0.034705.X_1$ $- 0.00267.X_2$ $- 0.00182.X_3$
	Intersection	-0.00267	0.000667	-4.00591	6.18E-05	
	Road hierarchy	0.034705	0.002084	16.65334	3.72E-62	
	Road width	-0.00182	0.000513	-3.54386	0.000395	
16 August	Intercept	0.264041	0.005713	46.21545	0	$Y = 0.264041 + 0.025617.X_1$ $- 0.00103.X_2$ $- 0.00105.X_3$
	Intersection	0.025617	0.001951	13.13185	2.39E-39	
	Road hierarchy	-0.00103	0.000622	-1.65167	0.098606	
	Road width	-0.00105	0.000478	-2.20189	0.027676	
17 August	Intercept	0.266207	0.00559	47.62615	0	$Y = 0.266207 + 0.024859.X_1$ $- 0.00165.X_2$ $- 0.00086.X_3$
	Intersection	0.024859	0.00191	13.01562	1.1E-38	
	Road hierarchy	-0.00165	0.000609	-2.71337	0.006662	
	Road width	-0.00086	0.000468	-1.84048	0.065702	

IV. CONCLUSION

This study demonstrated that traffic congestion in Samarinda City can be effectively identified and analyzed using a crowdsourcing approach that integrates real-time Google Maps data with spatial and statistical methods. The study shows how urban traffic is analyzed, based on the creation and testing of a new, integrated analytical framework. This framework takes advantage of the fact that Google Maps crowdsourced data are always available and economical. This makes it a better and more cost-effective option than traditional traffic monitoring systems. Methods like inductive loops, proprietary sensors, or dedicated GPS probe fleets are limited by high costs, ongoing maintenance needs, and most importantly, limited spatial resolution. This means that they can only give spot-speed data. This research, on the other hand, navigates around these challenges by providing complete network coverage and turning raw data into quantifiable, high-resolution metrics using advanced spatial and statistical

processing. The findings of this study demonstrate effective use of public data. It also involves making a scalable, immediately deployable, and financially commanding model that can be used in cities all over the world affected by the high costs of old hardware.

The hotspot analysis revealed that congestion was most prominent along major arterial and collector roads, particularly on the Samarinda–Bontang Road and Kebon Agung Road in the North Samarinda District. The regression analysis confirmed that intersections, road hierarchy, and road width are the key factors influencing congestion intensity in the city. Among these, intersections had the most consistent and significant impact across the observation period, highlighting the critical role of intersection management in improving traffic flow. Overall, integrating crowdsourced data with Geographic Information Systems (GIS) and statistical analysis proved to be a reliable, efficient, and cost-effective method for monitoring urban traffic dynamics. These findings provide policymakers

and urban planners with valuable insights to optimize road network design, prioritize intersection improvements, and implement data-driven strategies for sustainable transportation planning in Samarinda City.

The primary contribution lies in the advanced integration of specialized spatial and statistical methodologies, such as advanced hotspot detection and time-series analysis. Using these methods, the study turns crowdsourced data into statistically significant and actionable policy insights, providing analytical depth that is comparable to that of traditional expensive sensor networks. This established methodology is highly transferable and directly addresses critical urban planning challenges in resource-constrained environments globally, empowering rapidly urbanizing cities to transition decisively from anecdotal, reactive traffic management to data-driven optimization. The validated framework provides the information required to make objective, high-return infrastructure investment decisions, improve emergency service logistics, and optimize public transit routes.

Moreover, choosing Samarinda City as the study area supports the novelty and global relevance of the results. Samarinda is a complicated, non-Western place where growth happens quickly and often in an unplanned way. It also has a lot of traffic and unique planning challenges because it is close to Indonesia's new capital, IKN. By finding and measuring chronic congestion clusters and time patterns that are unique to the city's topography, the framework's efficacy and robustness have been successfully shown, providing the local government with objective and quantifiable evidence. The study provides an important example for how developing economies can use existing digital ecosystems to solve their complicated and growing mobility problems.

#### ACKNOWLEDGMENT

This research was funded by the Bima Regular Fundamental Research from the Ministry of Higher Education, Science, and Technology under contract No. 17620/IT10.L1/PPM.04/2025.

#### REFERENCES

- [1] Y. K. Thakre and P. Y. Pawade, "Traffic Congestion at Urban Road - Review," *IOP Conference Series: Earth and Environmental Science*, vol. 1326, no. 1, June 2024, Art. no. 012094, <https://doi.org/10.1088/1755-1315/1326/1/012094>.
- [2] D. Khoirotunni'mah, K. H. Putra, and T. M. C. Agusdini, "Kinerja Simpang Bersinyal dengan Metode PKJI 2023 dan MKJI 1997 pada Simpang Tiga Jalan Raya Mastrip, Wiyung, Surabaya," *Jurnal Talenta Sipil*, vol. 8, no. 1, Feb. 2025, Art. no. 50, <https://doi.org/10.33087/talentsipil.v8i1.784>.
- [3] J. Nagalapuram and S. Samundeeswari, "A Framework for Smart City Traffic Management utilizing BDA and IoT," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18989–18993, Dec. 2024, <https://doi.org/10.48084/etasr.8003>.
- [4] M. Z. Ibad, R. Sulistyorini, and C. Rahmah, "Google Traffic sebagai Masukan Kebijakan Transportasi Perkotaan (Studi Kasus: Kota Bandar Lampung)," *TATALOKA*, vol. 22, no. 3, pp. 409–417, Aug. 2020, <https://doi.org/10.14710/tataloka.22.3.409-417>.
- [5] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an Integrated Crowdsourcing Definition," *Journal of Information Science*, vol. 38, no. 2, pp. 189–200, Apr. 2012, <https://doi.org/10.1177/0165551512437638>.
- [6] G. D. Saxton, O. Oh, and R. Kishore, "Rules of Crowdsourcing: Models, Issues, and Systems of Control," *Information Systems Management*, vol. 30, no. 1, pp. 2–20, Jan. 2013, <https://doi.org/10.1080/10580530.2013.739883>.
- [7] Sugiyono, *Metode Penelitian Bisnis : Pendekatan Kuantitatif, Kualitatif, Kombinasi dan R&D*. Bandung, Indonesia: Alfabeta, 2013.
- [8] X. Li *et al.*, "Development of Geographic Information System Architecture Feature Analysis and Evolution Trend Research," *Sustainability*, vol. 16, no. 1, Dec. 2023, Art. no. 137, <https://doi.org/10.3390/su16010137>.
- [9] M. A. Simanungkalit, A. Hidayat, R. A. Nugroho, and A. S. Depari, "Analisis Hotspot (Getis Ord Gi\*) Pola Spasial Frekuensi Kecelakaan Lalu Lintas di Kota Balikpapan," *COMPACT: Spatial Development Journal*, vol. 3, no. 1, June 2024, <https://doi.org/10.35718/compact.v3i1.1156>.
- [10] "Samarinda City," *Google Maps*, Aug. 2025. <https://www.google.com/maps/place/Samarinda+City,+East+Kalimantan,+Indonesia/> (accessed Aug. 11, 2025).
- [11] "OpenStreetMap-East Kalimantan (5449459)," *OpenStreetMap*, Aug. 2025. <https://www.openstreetmap.org/relation/14921097> (accessed Aug. 11, 2025).