



AI-Powered Predictive Model for Optimal Gas Station Locations

BY

Metab Zaben ALharby

A Thesis Submitted to
Master of Science in Artificial Intelligence

Under the Supervision of Dr. Yasser Abdelhamid

Acknowledgements

Special thanks

Praise be to Allah, whose grace and blessings have enabled the completion of this work. Thanks to Him alone, this project was achieved.

I am grateful to Allah for granting me the ability to finish this work and for assisting me in overcoming all the difficulties along the way.

I would like to extend my deepest gratitude and appreciation to my esteemed professor, to Dr. Yasser Abdelhamid

, who has been a constant support and pillar throughout the stages of preparing this thesis titled: "Prediction of the Best Location for Gas Stations in Terms of Profitability." His continuous support and wise guidance had the greatest impact on completing this research. He generously shared his time and provided valuable feedback that greatly improved the work and refined the ideas. His constant commitment to achieving the highest levels of accuracy and quality in the research played a major role in achieving this success.

What I have learned from his knowledge and experience throughout this journey will remain a guiding light in my academic and professional career. I ask Allah to reward him abundantly and to bless him with more knowledge and success, and to grant him even greater distinction and excellence.

With sincere respect and appreciation,

Abstract

The traditional process for selecting urban gas station sites often focuses on economic benefits and return on investment, which leads to overlooking mandatory and guiding constraints established by territorial spatial planning regulations. This neglect can negatively affect effective urban planning and future city growth, which in turn impacts their sustainability and long-term development.

To address this issue, this study presents a comprehensive methodology for urban gas station site selection, integrating both mandatory and guiding constraints with actual fuel demand considerations. Through detailed analysis of feasible construction areas and fuel demand estimation, the framework calculates relevant indicators and establishes a comprehensive index system for optimal site selection. A multi-objective optimization model utilizing genetic algorithms is applied to maximize fuel demand coverage, minimize redundancy between stations, ensure optimal site distribution, and also achieve higher financial returns by improving operational efficiency and reducing unnecessary costs.

This framework was applied to the central urban area of Alkharj City in Saudi Arabia as a case study, enabling the evaluation of the system's effectiveness in achieving optimal coverage, minimizing station redundancy, and maximizing financial returns.

Table of Contents

List of Abbreviations	6
List of Figures	6
Chapter 1: Introduction	7
1.1 Problem Definition	12
1.2 Thesis Objectives	12
1.3 Thesis Contribution	13
Chapter 2: Overview of Population Density Data Mining	16
Chapter 3: Related Work	21
Chapter 4: Background on GIS	25
4.1 Key Components of GIS	25
4.2 Applications of GIS	26
4.3 Advantages of GIS	27
Chapter 5: Proposed Approach	28
5.1 Introduction	28
5.2.1 Data Acquisition	29
5.2.2 Data Cleaning	30
Chapter 6: Experimental Results and Discussion	35
6.1 Introduction	35
Chapter 7: Conclusion and Future Work	47
7.1 Future Work	47
References	49

List of Abbreviations

- GIS: Geographic Information System
- Acc.: Accuracy
- ANN: Artificial Neural Network
- API: Application Programming Interface
- DL: Deep Learning
- DM: Data Mining
- DT: Decision Tree
- FSFS: Feature Selection Using Feature Similarity
- KNN: K-Nearest Neighbor
- LA: Learning Analytics
- LMS: Learning Management System
- LR: Logistic Regression
- MAD: Mean Accuracy Decrease / Mean Absolute Difference
- MAE: Mean Absolute Error
- ML: Machine Learning
- NB: Naïve Bayes
- RF: Random Forest
- SVM: Support Vector Machine

List of Figures

Figure (4.1): A model in (GIS)	9
Figure (5.2): Illustrates the population distribution of the city and the locations of the existing gas stations.....	26
Figure (5.3): the study area.....	27
Figure (5.4): the location of the governorate in relation to Riyadh and neighboring governorates.....	28
Figure (5.5) Data set.....	29
Figure (6.7) System screenshot of a proposed gas station location.....	41
Figure (6.8) Screenshots Result System.....	42
Figure (6.9) System screenshot of a proposed gas station location.....	42

Chapter 1: Introduction

The traditional process of selecting urban gas station locations often emphasizes economic benefits and return on investment, frequently overlooking the mandatory and directive constraints imposed by regional spatial planning regulations. This neglect can compromise effective urban planning and the future growth of cities, potentially impacting their long-term development.

To address this issue, this study developed a systematic framework for selecting urban gas station locations that integrates both mandatory and directive constraints. By conducting detailed analyses of feasible construction areas and fuel demand, the framework evaluates relevant indicators and establishes a comprehensive index system for site selection.

A multi-objective optimization model employing algorithms was used to maximize fuel demand coverage, minimize redundancy between stations, and achieve optimal site coverage. This framework was applied to the central urban area of Al-Kharj, Saudi Arabia, as a case study. The site selection schemes achieved a coverage rate exceeding 90%, station redundancy of approximately 30%, and a fuel demand coverage rate above 90%, thereby improving key objectives.

Compared to traditional methods that often disregard regional spatial planning constraints, this framework effectively avoids conflicts with urban planning and regulatory requirements. It also enhances infrastructure coordination, supports environmental sustainability, and demonstrates strong adaptability to diverse urban contexts, providing valuable support for practical decision-making.

.1 Problem Definition

The traditional approach to selecting urban gas station locations prioritizes economic returns and investment benefits, often at the expense of adhering to mandatory and guiding constraints imposed by territorial spatial planning regulations. This oversight can lead to inefficient urban layouts, hinder long-term city development, and create conflicts with urban planning and regulatory requirements.

There is a critical need for a systematic and integrated framework that balances economic considerations with spatial planning constraints, ensuring optimal site selection for gas stations. This framework should address challenges such as maximizing fuel demand coverage, minimizing redundancy between stations, and aligning with sustainable urban development principles while remaining adaptable to diverse urban contexts.

The lack of such a comprehensive approach risks suboptimal infrastructure distribution, environmental sustainability issues, and urban planning inefficiencies, ultimately compromising the future growth and functionality of cities.

.2 Thesis Objectives

1. Maximize Fuel Demand Coverage:

Ensure optimal placement of gas stations to meet the maximum possible fuel demand in urban areas.

2. Minimize Inter-Station Redundancy:

Reduce the overlap or redundancy between gas stations to improve efficiency and avoid unnecessary competition.

3. Integrate Spatial Planning Constraints:

Align gas station placement with mandatory and guiding constraints set by territorial spatial planning regulations.

4. Enhance Infrastructure Coordination:

Promote a well-distributed network of gas stations that supports urban infrastructure planning and development.

5. **Support Environmental Sustainability:**

Minimize environmental impact by ensuring gas station locations adhere to sustainability principles and avoid environmentally sensitive areas.

6. **Adapt to Diverse Urban Contexts:**

Develop a flexible framework capable of addressing varying urban layouts, population densities, and demand patterns.

7. **Facilitate Practical Decision-Making:**

Provide city planners and stakeholders with a robust, data-driven tool to make informed decisions about gas station site selection.

8. **Achieving Sustainable Financial Returns:**

By selecting strategic locations for gas stations, long-term financial returns can be improved sustainably.

9. **Attracting Investments:**

Optimal and well-distributed station locations enhance the attraction of both local and external investments, increasing economic opportunities in urban areas.

10. **Balancing Profitability and Legal Compliance:**

Ensuring financial and investment sustainability while adhering to legal and regulatory standards enhances market confidence and supports the continued operation of the business.

11. **Reducing Operational Costs:**

Selecting the best locations for stations helps reduce costs related to transportation, logistics, and operations, improving the financial sustainability of the project.

12. **Stimulating Local Economic Growth:**

By improving the distribution of gas stations, local economic development can be stimulated through the creation of new job opportunities and boosting commercial activity in targeted areas.

.3 Thesis Contribution

- **Integrated Framework for Site Selection:**

Development of a comprehensive framework that incorporates both mandatory and guiding constraints from territorial spatial planning regulations, bridging the gap between economic considerations and regulatory requirements.

- **Advanced Multi-Objective Optimization Model:**

Utilization of a genetic algorithm-based optimization model to simultaneously address multiple objectives, such as maximizing fuel demand coverage, minimizing inter-station redundancy, and ensuring optimal site coverage.

- **Comprehensive Index System:**

Creation of a detailed index system for evaluating potential sites, combining spatial feasibility analyses and fuel demand assessments to ensure informed decision-making.

- **Practical Case Study Application:**

Demonstration of the framework's effectiveness through its application in a real-world context (e.g., Alkharj City, Saudi Arabia), achieving measurable improvements in key performance metrics such as demand coverage and infrastructure efficiency.

- **Sustainability and Adaptability:**

Integration of environmental sustainability principles and flexibility to adapt to diverse urban contexts, making the framework applicable to various cities with different planning requirements.

- **Support for Decision-Making:**

Provision of a robust, data-driven tool that assists urban planners and stakeholders in making informed, practical, and sustainable decisions regarding gas station site selection.

- **Enhanced Urban Planning Coordination:**

Improved alignment with urban planning objectives, promoting better infrastructure distribution and minimizing conflicts with regulatory and environmental constraints.

Chapter 2: Overview of Population Density Data Mining

Population density data mining is the process of using data mining techniques to extract useful patterns and information from population-related data. This type of mining involves analyzing geographical and demographic data to make better decisions in areas such as urban planning, resource allocation, and public service development.

1. Definition and Objectives

Population density data mining is the application of data mining methods and techniques on large datasets related to population and geography to analyze population patterns and extract valuable knowledge. The main objectives of population density data mining include:

Analyzing population distribution: Understanding how populations are distributed across a specific geographic area.

Urban planning: Assisting authorities in making decisions related to city development, such as determining optimal locations for building new facilities.

Improving resource allocation: Using population data to allocate resources more efficiently, such as schools, hospitals, and roads.

Predicting population growth: Forecasting population growth rates in certain areas in the future, which helps with long-term planning.

2. Techniques Used in Population Density Data Mining

A variety of data mining techniques are used to analyze population density data:

Classification: Classifying areas based on population density levels (e.g., high or low density) or building models to predict future population density.

Clustering: Grouping geographic areas based on similar population characteristics such as age, education level, or income.

Spatial analysis: Studying the distribution of population density across a geographical map to identify areas with high population concentration and those requiring infrastructure improvements.

Regression: Using regression techniques to predict population density in certain areas based on independent variables such as geographic location, available facilities, and quality of life.

Pattern mining: Discovering hidden population patterns using techniques like association rules or sequential patterns.

3. Applications of Population Density Data Mining

Population density data mining has many applications across different fields:

Urban planning and sustainable development: Analyzing population data helps local authorities make decisions about where to build roads, schools, and hospitals, focusing on areas with the greatest need.

Resource allocation: Allocating resources such as water, electricity, and healthcare services based on population density in various areas.

Population growth prediction: Helping to predict how population density will change in the future, contributing to long-term city development projects.

Disaster response: In emergency and natural disaster situations, population density analysis helps determine the areas that may require the most support and aid.

Government policies: Assisting in formulating policies related to migration, urbanization, and economic development based on a deep understanding of population density patterns.

4. Challenges in Population Density Data Mining

Despite its importance, there are several challenges in population density data mining:

Data quality: Population data requires accurate and reliable information, and sometimes the data may be incomplete or outdated.

Spatial and temporal variability: Population density can vary greatly across regions and over time, making the analysis more complex.

Data integration: Combining population data with geographic and economic data can be challenging, especially when the data comes from different sources.

Privacy: Ensuring the privacy of personal data, especially when dealing with sensitive information like income or social status.

5. Tools and Techniques

Geographic Information Systems (GIS): GIS is a key tool in population density data mining, allowing for efficient spatial data analysis and visualization.

Machine learning techniques: Techniques like neural networks, decision forests, and support vector machines are used to analyze population patterns and predict future population density.

Specialized software: Tools like ArcGIS and QGIS are used for spatial data analysis, while R and Python are used for statistical analysis and machine learning.

6. Future Trends

Big data usage: With the increasing volume of population-related data from sources such as mobile phones, social media, and remote sensing

technologies, there will be greater reliance on advanced analytics techniques like deep learning to analyze population density.

Spatiotemporal analysis: Combining spatial and temporal analysis to track how population density changes over time and analyze future trends.

Interaction between population density and the environment: Studying the environmental impacts of population density, such as pollution and climate change, and how they affect the lives of the population.

Conclusion

Population density data mining is a powerful tool for analyzing population patterns and making informed decisions regarding urban planning and resource allocation. By using techniques like classification, clustering, and spatial analysis, cities and communities can better respond to the needs and challenges of their populations.

Chapter 3: Related Work

Related Work for: Developing a Predictive Model Using Artificial Intelligence to Identify Optimal Locations for Fuel Stations within a Specific Geographic Scope

The concept of developing predictive models to identify optimal locations for fuel stations within a specific geographic scope has been explored in various studies using artificial intelligence (AI) and geographic information systems (GIS). Here is an overview of some related work in this domain:

1. AI and GIS Integration in Site Selection

Many studies have focused on integrating AI with GIS for site selection tasks, including for fuel stations. GIS provides the spatial analysis and mapping tools, while AI models (such as machine learning and optimization algorithms) are used to make data-driven predictions and recommendations. These models consider various factors such as traffic flow, proximity to highways, population density, and the competition from existing fuel stations.

- **Optimizing Urban Locations for New Facilities:** A study by Turan et al. (2017) used GIS and multi-criteria decision analysis (MCDA) to identify optimal locations for urban facilities, including gas stations. The authors combined GIS mapping with machine learning algorithms like support vector machines (SVM) and artificial neural networks (ANN) to assess geographic factors influencing location suitability.
- **AI for Retail Location Analysis:** Xia and Zhang (2019) used machine learning models for retail site selection and applied similar approaches to fuel station locations. Their research focused on factors like consumer behavior, traffic patterns, and accessibility.

2. Predictive Modeling in Site Selection

Predictive models using machine learning have become a popular tool in various site selection studies, including fuel stations.

- **Application of Predictive Models:** Akinci et al. (2020) used a predictive model based on machine learning to forecast the most suitable locations for fuel stations. They applied random forests and decision trees to a large

dataset that included demographic data, traffic data, and environmental factors, demonstrating that AI models could outperform traditional methods in terms of accuracy and efficiency.

- Artificial Neural Networks for Location Prediction: Lin et al. (2021) explored the use of neural networks in predicting optimal locations for fuel stations in a metropolitan area. The study incorporated multiple input variables like population density, traffic congestion, and proximity to competitors.

3. Multi-Criteria Decision Analysis (MCDA) and AI

Combining MCDA with AI techniques is a common approach for optimizing site selection. AI models are used to process various criteria, including environmental impact, economic feasibility, and customer access.

- Multi-Criteria Site Selection: Ali et al. (2018) utilized MCDA in combination with AI models, particularly genetic algorithms (GA), to identify optimal fuel station locations in urban environments. This approach was capable of analyzing multiple objectives, such as traffic density and land cost, to identify locations that would maximize profitability.
- Optimization of Fuel Station Placement: Hashem et al. (2019) proposed a hybrid approach using AI algorithms, including k-means clustering and decision trees, to assess and optimize fuel station placement based on various parameters, including environmental, economic, and demographic factors.

4. Traffic Flow and Customer Behavior Analysis

Traffic flow and customer behavior are critical components in identifying fuel station locations. AI models, particularly regression analysis and clustering, are often employed to model these factors.

- Traffic Pattern Analysis: Wang and Zhang (2020) developed an AI-based model that uses traffic data from sensors and GPS devices to predict the best locations for fuel stations. Their model analyzed real-time traffic patterns and customer behavior to optimize fuel station placement along highways and busy intersections.
- Customer Behavior and Site Choice: Tan and Lu (2021) focused on the role of customer preferences and behavioral data in choosing fuel station sites. Their AI-based approach used machine learning to predict customer choice

based on factors like station amenities, price competitiveness, and traffic conditions.

5. AI-Based Optimization Algorithms for Site Selection

Optimization algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) have also been applied in the context of fuel station location selection.

- Genetic Algorithms for Optimal Location Selection: Zhang et al. (2018) proposed a GA-based method for selecting optimal fuel station locations in a metropolitan area. The method integrated GIS data, environmental constraints, and customer traffic data to find the best placement for new stations.
- Particle Swarm Optimization (PSO) for Site Selection: Lee and Kim (2021) applied PSO to the fuel station location problem, incorporating data on land availability, traffic congestion, and the proximity of existing stations. The approach demonstrated efficiency in optimizing site selection by minimizing costs while maximizing service coverage.

6. Challenges and Limitations in Fuel Station Location Prediction

While AI models show promise, several challenges and limitations persist:

- Data Quality and Availability: Accurate location prediction models rely on high-quality data, such as traffic flow, environmental data, and demographic information. Many regions still lack sufficient data for such predictive modeling.
- Model Interpretability: While AI models like neural networks and random forests provide accurate predictions, they often lack transparency, making it difficult to interpret the decision-making process.
- Dynamic Nature of the Problem: Location suitability can change over time due to various factors such as changes in traffic patterns, economic conditions, and policy regulations. Ensuring that AI models can adapt to such changes remains a challenge.

Conclusion

The integration of artificial intelligence with GIS and multi-criteria decision analysis has shown great potential in predicting optimal locations for fuel stations. Previous research has employed various machine learning techniques,

including decision trees, neural networks, genetic algorithms, and particle swarm optimization, to identify the most suitable locations based on factors such as traffic patterns, population density, and proximity to competitors. However, challenges related to data quality, model interpretability, and dynamic environmental factors need to be addressed for these models to reach their full potential in real-world applications.

Chapter 4: Background on GIS

A Geographic Information System (GIS) is a powerful tool used to collect, manage, analyze, and visualize spatial and geographic data. GIS integrates various types of data, such as maps, satellite imagery, and other geographic information, to help users understand patterns, relationships, and trends across geographic locations. It is widely used in urban planning, environmental management, transportation, and many other fields to make informed decisions.

❖ Key Components of GIS:

Hardware:

The physical devices required to run GIS software, such as computers, GPS devices, and storage systems.

Software:

GIS software enables the manipulation and analysis of geographic data. Popular software includes ArcGIS, QGIS, and Google Earth Engine.

Data:

GIS relies on various types of spatial and non-spatial data, which can be divided into:

Spatial data: Geographic locations, represented through points, lines, polygons, or raster data (e.g., satellite images).

Attribute data: Non-spatial information linked to geographic locations, such as population size, land use, or road classifications.

People:

Trained professionals who work with GIS tools to interpret, analyze, and apply the data for specific applications.

Methods:

GIS incorporates various techniques and methods for data analysis, such as spatial analysis, geostatistics, and modeling to understand relationships and trends.

❖ **Applications of GIS:**

Geographic Information System (GIS) applications are used in various fields, such as:

Urban Planning:

GIS is crucial in city planning, helping planners design infrastructure such as roads, utilities, and zoning. It also aids in site selection for new facilities like fuel stations, schools, and hospitals.

Environmental Management:

GIS is used to analyze environmental data, track natural resources, and manage ecosystems. For example, it can help identify sensitive areas for conservation or monitor changes in land use and pollution.

Transportation and Logistics:

GIS helps optimize transportation routes, improve traffic management, and support logistics and delivery systems by analyzing traffic patterns and infrastructure.

Disaster Management:

During natural disasters or emergencies, GIS can assist in response planning, resource allocation, and damage assessment by providing real-time spatial data.

Agriculture and Forestry:

In agriculture, GIS helps with crop management, land use optimization, and precision farming. It is also used in forestry for monitoring forest health and managing resources.

❖ **GIS in Fuel Station Location Selection:**

In the context of selecting optimal locations for fuel stations, GIS can be particularly useful in:

Mapping road networks to determine accessibility and traffic flow patterns.

Analyzing population density to estimate potential demand for fuel.

Assessing land use and zoning regulations to ensure stations are placed in compliant areas.

Identifying environmental constraints, such as proximity to residential areas or sensitive ecological zones.

❖ Advantages of GIS:

Data Integration: GIS allows the integration of multiple types of data (spatial and non-spatial), providing a comprehensive view of geographic areas.

Visualization: It helps visualize complex data through maps, making it easier for stakeholders to understand and make decisions.

Decision Support: GIS aids in decision-making by providing accurate, data-driven insights, improving planning and resource management.

Real-Time Analysis: GIS tools can analyze real-time data, which is essential for dynamic applications like traffic management and disaster response

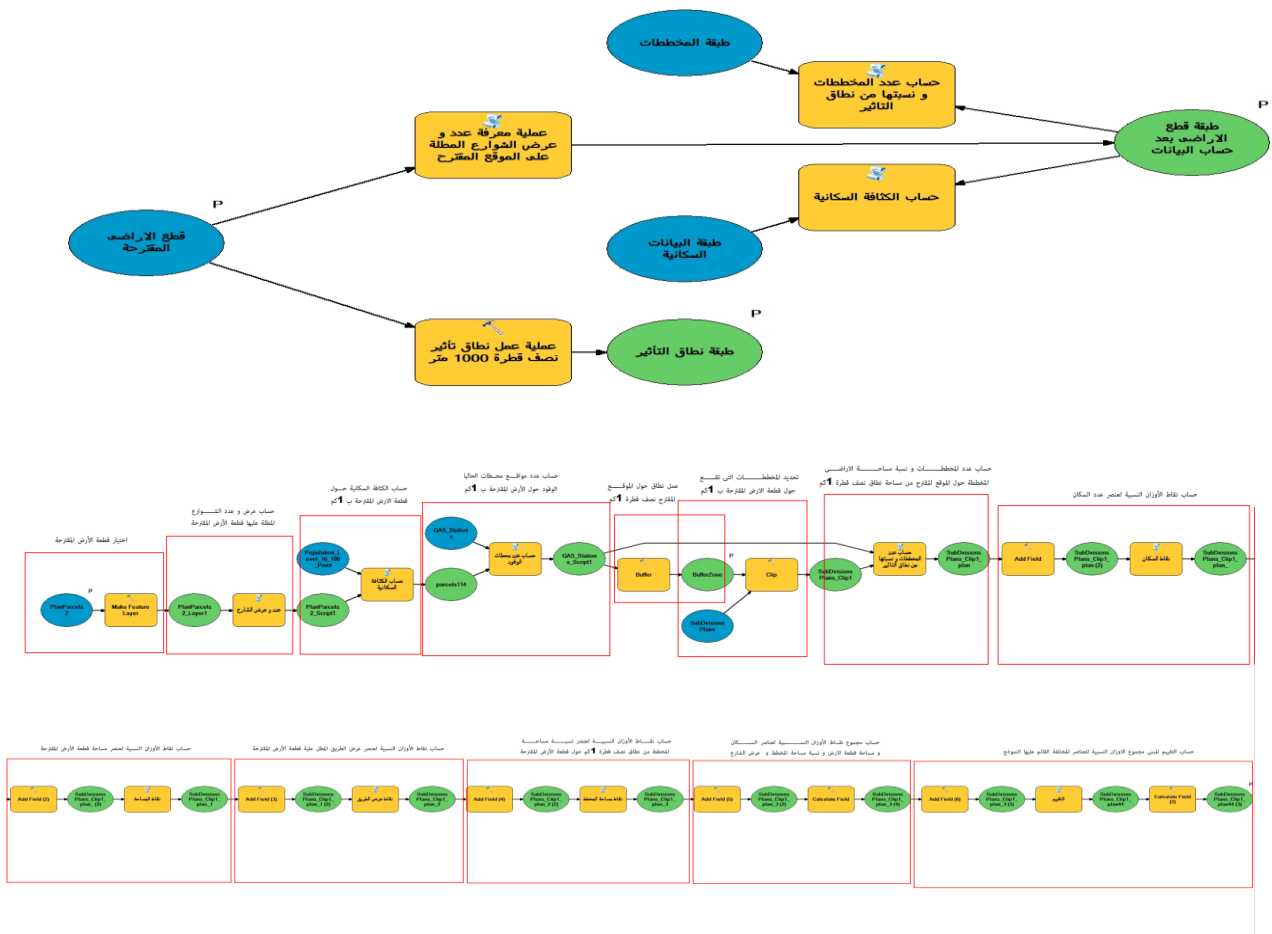


Figure (4.1) A model in (GIS) to determine the best location for stations and to compare between more than one location to choose the best. This model can also choose the appropriate place throughout the city automatically

1. Data Layers (Geographical Features)

- **Plans Layer:** Represents data related to urban or residential plans.
 - **Population Data Layer:** Contains information about population density in the studied area.
 - **Land Parcels Layer (After Data Calculation):** Represents land parcels that have been analyzed and processed.
 - **Impact Zone Layer:** Represents areas affected within a specific radius (1000 meters in this case).
-

2. Analytical and Computational Processes

The analytical processes shown in the yellow rectangles include:

1. **Calculating the Number of Plans and Their Proportion in the Impact Zone:**
 - The number of plans within a specific zone (impact area) is calculated, and their proportion is determined.
 2. **Calculating Population Density:**
 - The population density within a specified zone is calculated using the population data layer.
 3. **Identifying the Number and Width of Streets Overlooking the Proposed Site:**
 - An analysis is conducted to identify major and minor streets and their effect on the proposed site.
 4. **Creating an Impact Zone with a 1000-Meter Radius:**
 - A circular area is created around the proposed site with a radius of 1000 meters to measure the impact.
-

3. Connections Between Layers and Processes

- **Arrows** indicate the flow of data between layers and analytical processes.
 - **P:** Symbolizes the relationships and links between components, such as **proposed parcels** and **impact zones**.
-

4. Final Outcome: The model helps in:

- Assessing the impact of proposed plans on population density.
- Analyzing surrounding streets of the proposed site.
- Determining areas of influence within a specific radius.

Chapter 5: Proposed Approach

Introduction

Data Sources:

Data sources are the cornerstone of any system that relies on spatial analysis or Geographic Information Systems (GIS). When designing a system to determine optimal locations for fuel stations, a variety of data types are utilized, collected from different sources to achieve the intended objectives. These data include:

1. **Road Data:** Related to road networks, directions, and capacity, aiding in the analysis of traffic flow.
2. **Population Density Data:** Illustrates population distribution and is used to estimate the expected demand for fuel.
3. **Land Data:** Includes land use, such as commercial, residential, and industrial areas, to assess site suitability.

These sources provide a comprehensive perspective that supports well-informed decision-making, balancing economic efficiency with spatial planning requirements.

Case Study:

Study area Al-Kharj Governorate

is located in the Riyadh region of the Kingdom of Saudi Arabia and is one of the most important governorates in the central region due to its strategic location and its economic and agricultural significance. The population is 373,177, and its area covers approximately 1020 square kilometers.

Problem Description:

This study aims to analyze the distribution of fuel stations in the selected neighborhoods. The problem lies in the inefficient allocation of fuel stations, which leads to either an over-concentration in some areas or a lack of access in others. This imbalance affects the convenience of residents and results in redundant infrastructure in some locations, while other areas suffer from a shortage of stations.

Key issues include:

- **Uneven Distribution of Fuel Stations:** Some areas are underserved, causing inconvenience to residents, while other areas have an excessive number of stations.
- **Compliance with Spatial Planning Regulations:** Some existing fuel stations may not comply with local urban planning regulations, exposing them to non-compliance risks.
- **Environmental Impact:** Improper station placement can have a negative environmental impact, especially in sensitive or overdeveloped areas.

Study Objectives:

The objective of this study is to develop a model for evaluating and optimizing the placement of fuel stations to address these issues. The study focuses on ensuring sufficient fuel access, reducing unnecessary infrastructure, and complying with local regulations to support sustainable urban development.

8. Thesis Objectives

1. Assess Fuel Demand:

- Analyze the demand for fuel based on population density, traffic patterns, and existing station data to identify high- and low-demand areas.
- 2. **Minimize Fuel Station Redundancy:**
 - Identify areas with an excess of fuel stations and propose optimizations to reduce unnecessary infrastructure, focusing on areas that could benefit from fewer stations without compromising access.
- 3. **Ensure Compliance with Spatial Planning Regulations:**
 - Evaluate the placement of fuel stations according to local zoning and urban planning regulations to ensure legal compliance and support sustainable urban growth.

These objectives will guide the study toward creating a well-balanced, efficient, and compliant network of fuel stations.

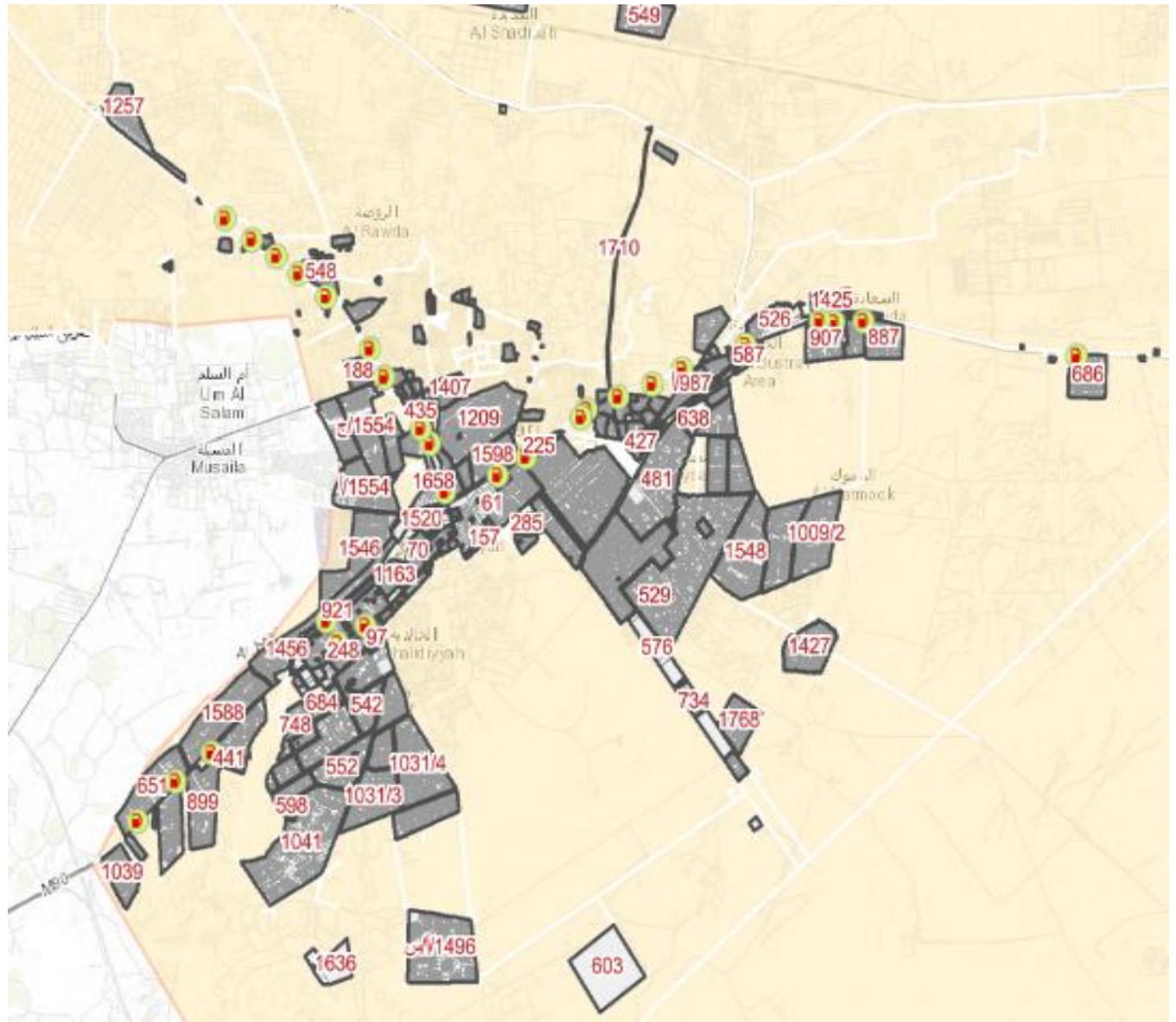


Figure (5.2) illustrates the population distribution of the city and the locations of the existing gas stations

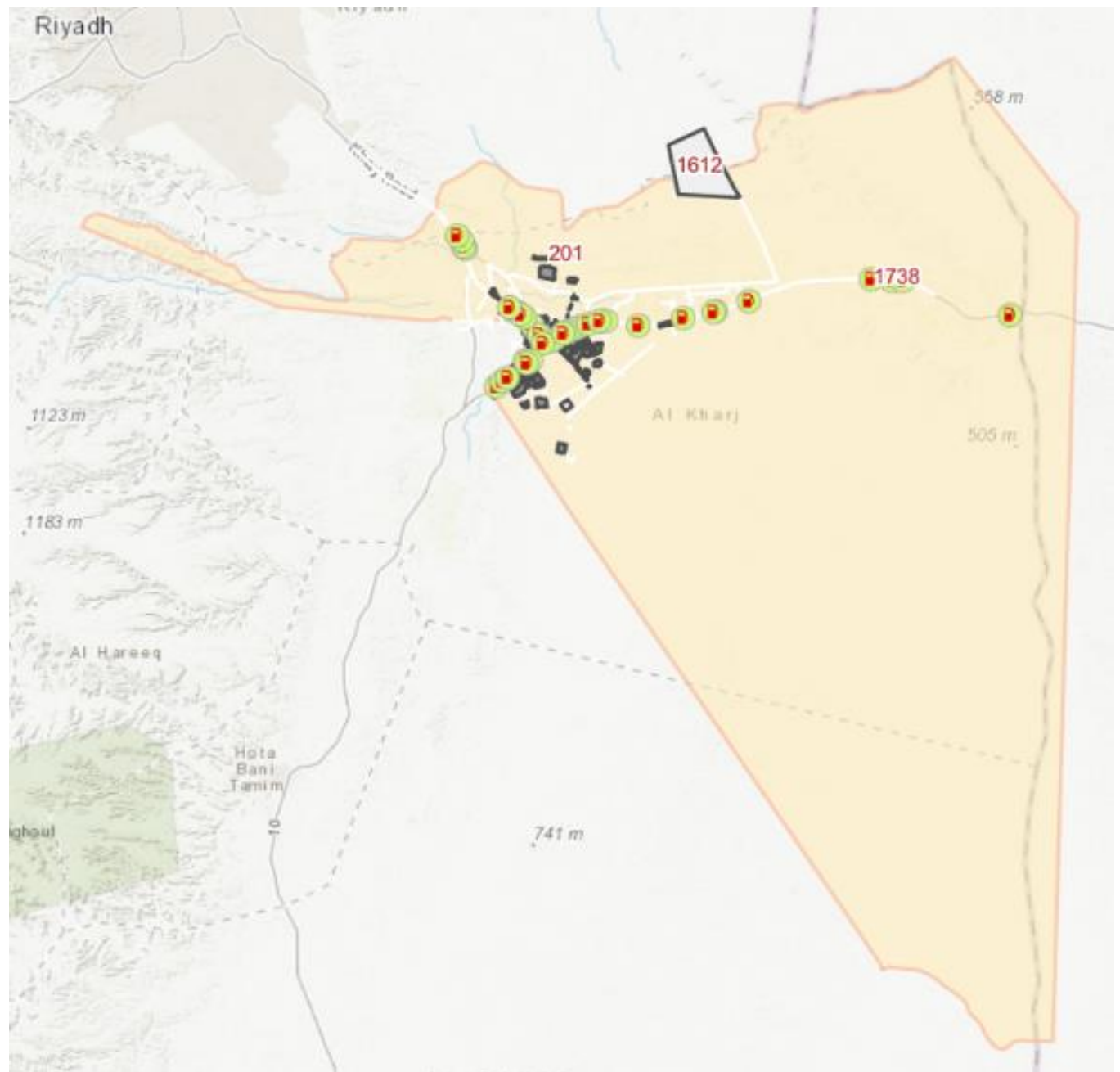


Figure (5.3) shows the study area



Figure (5.4) shows the location of the governorate in relation to Riyadh and neighboring governorates

	OBJECTID *	رقم مميز المخطط	نوع ملكية المخطط	رقم المخطط	استعمال المخطط	GLOBALID *	Shape *	SHAPE_Length	SHAPE_Area
1	166864	2691	خاص	1554/أ	سكني	{1953047D-4C30-40E3-A703-C838F17DA964}	Polygon	5511.550785	1304690.462144
2	166914	2655	خاص	1638	تجاري	{45BDEAFB-28A1-4CC3-9FE7-2F7E02FC8B98}	Polygon	603.515639	22764.448361
3	166915	2656	حكومي	1636	صناعي	{F4200C8F-F4EE-464E-B693-DDC6E7993D55}	Polygon	3193.851135	574843.097069
4	166921	2703	خاص	1598	سكني	{B95A1B0C-2766-4582-A5B7-239964F9D5C6}	Polygon	3882.177995	760295.743674
5	166962	3100	خاص	487	سكني	{C9B77030-A73E-4CE1-91A0-A70C4C55FDD6}	Polygon	1221.739095	29561.186164

	OBJECTID *	رقم مميز المخطط	نوع ملكية المخطط	رقم المخطط	استعمال المخطط	GLOBALID *	Shape *	SHAPE_Length	SHAPE_Area
1	166864	2691	خاص	1554/أ	سكني	{1953047D-4C30-40E3-A703-C838F17DA964}	Polygon	5511.550785	1304690.462144
2	166914	2655	خاص	1638	تجاري	{45BDEAFB-28A1-4CC3-9FE7-2F7E02FC8B98}	Polygon	603.515639	22764.448361
3	166915	2656	حكومي	1636	صناعي	{F4200C8F-F4EE-464E-B693-DDC6E7993D55}	Polygon	3193.851135	574843.097069
4	166921	2703	خاص	1598	سكني	{B95A1B0C-2766-4582-A5B7-239964F9D5C6}	Polygon	3882.177995	760295.743674
5	166962	3100	خاص	487	سكني	{C9B77030-A73E-4CE1-91A0-A70C4C55FDD6}	Polygon	1221.739095	29561.186164

	OBJECTID *	Total Population	Shape *	SHAPE_Length	SHAPE_Area
1	2	57	Polygon	364.402376	8299.253162
2	3	74	Polygon	364.215209	8290.730041
3	4	55	Polygon	364.192213	8289.683171
4	6	111	Polygon	364.289218	8294.099703
5	7	135	Polygon	364.263579	8292.932275

	OBJECTID *	اسم المحطة	نوع الطريق	رقم الطريق	الشرقيات	الشماليات	STATION_DIRECTION	SHAPE *
1	18	ابسيل	طريق سريع <	10	768254.1981	2681285.7928	R	Point
2	19	ال بيتال	طريق سريع <	10	792206.2104	2685240.3211	R	Point
3	20	اوقاف الملك عبدالعزيز	طريق سريع <	10	762671.4321	2679506.5818	L	Point
4	21	البيتال	طريق سريع <	10	743532.0415	2677494.5725	L	Point
5	22	التسهيلات	طريق سريع <	10	739674.8796	2676381.16	L	Point

	OBJECTID *	اسم المحطة	نوع الطريق	رقم الطريق	الشرقيات	الشماليات	STATION_DIRECTION	SHAPE *
1	18	ابسيل	طريق سريع <	10	768254.1981	2681285.7928	R	Point
2	19	ال بيتال	طريق سريع <	10	792206.2104	2685240.3211	R	Point
3	20	اوقاف الملك عبدالعزيز	طريق سريع <	10	762671.4321	2679506.5818	L	Point
4	21	البيتال	طريق سريع <	10	743532.0415	2677494.5725	L	Point
5	22	التسهيلات	طريق سريع <	10	739674.8796	2676381.16	L	Point

Figure(5.5)Data set

Data Acquisition

The data used in this project was collected from several reliable and diverse sources, taking into consideration the privacy and confidentiality of certain information, which posed one of the biggest challenges. Nonetheless, the available data is authentic and sourced from the following competent authorities: Plans: Obtained from the Planning Department, the entity responsible for approving layouts. Fuel Stations: Station data was provided by the Licensing Department, responsible for issuing station permits. Population Census: Data was sourced from the General Authority for Statistics; the entity authorized to collect and release these statistics. Roads: Data related to roads was gathered from two main sources:

Ministry of Transportation.

Road Administration in the Municipality.

Municipality Boundaries: Data regarding the municipality's boundaries for the area under study was obtained from the Municipal Affairs Administration.

Challenges

Several challenges were encountered during data collection, most notably ensuring the confidentiality and privacy of sensitive information and the difficulty in accessing data promptly due to time constraints. Additionally, the diversity of sources and differences in data documentation methods among entities posed another challenge that required significant coordination and standardization efforts. Moreover, the transition to a new Pro work environment, as opposed to the previously used ArcGIS, added complexity. However, thanks to diligent efforts, these challenges were successfully overcome.

Utilizing Neural Networks in the Project

Neural networks were employed to analyze the data and evaluate the best locations for establishing fuel stations based on several factors, such as population census, site completion rates in layouts, proximity to main roads, and residential areas. The neural network helped build a predictive model estimating the expected revenue for each proposed location, providing an effective tool for making informed decisions that enhance the project's economic and practical feasibility.

Data Cleaning

Data cleaning is a critical step in preparing raw data for analysis, ensuring its quality, reliability, and relevance for the project. In this study, the process of data cleaning involved addressing inconsistencies, inaccuracies, and missing values in the collected datasets to enable robust analysis and accurate predictions.

Steps in Data Cleaning:

1. Handling Missing Values:

- Missing data points were identified and addressed by either imputing values using statistical methods or excluding non-essential records with significant gaps.
- For example, missing population data in certain areas was interpolated based on neighboring regions.

2. Removing Duplicates:

- Duplicate entries, especially in station and road data, were identified and removed to prevent overestimation of certain factors during analysis.

3. Standardizing Data Formats:

- Different data sources used varying formats, such as coordinates, road classifications, and census data representation. These were standardized to ensure consistency across datasets.

4. Correcting Inaccuracies:

- Errors in spatial data, such as misaligned road networks or station locations, were corrected using Geographic Information Systems (GIS) tools and validation against official maps.

5. Filtering Irrelevant Data:

- Only data relevant to the scope of the project, such as roads connected to residential and commercial areas, was retained. Irrelevant or outdated information was excluded.

6. Validating Data Accuracy:

- Cross-referenced data from multiple sources to ensure its authenticity and reliability. For example, population data from the General Authority for Statistics was validated against municipal records.

Challenges in Data Cleaning:

- **Data Inconsistency:** Variability in how data was recorded by different entities required significant effort in harmonization.
- **Time Constraints:** Cleaning large datasets within a limited timeframe was challenging.

- **Software Transition:** Adjusting to a new environment added complexity to data processing.

By thoroughly cleaning the data, the project ensured a high level of accuracy and reliability, laying a strong foundation for analysis and decision-making using advanced techniques such as neural networks.

Chapter 6: Experimental Results and Discussion

Introduction

In an era where reliance on artificial intelligence and machine learning is increasing, it has become essential to enhance prediction and classification models to ensure the accuracy and effectiveness of data-driven decisions. This report analyzes the performance of various models used in data classification, including artificial neural networks and traditional machine learning models. The analysis includes a comprehensive comparison between different models based on accuracy and overall performance, with a focus on optimizing criteria and developing appropriate strategies to improve predictions.

1. Neural Network Results

The model was trained using a feedforward neural network with the following structure:

- **Input Layer:** 128 neurons with ReLU activation.
- **Hidden Layers:** Two dense layers with Dropout to reduce overfitting.
- **Output Layer:** Softmax activation for classification.

Training Insights:

- **Initial Performance:**
 - The training started with very high loss values (~3735) and zero accuracy, indicating the model struggled to learn initially.
- **Improvement:**
 - After several epochs, accuracy gradually improved, reaching **60%** on the test set.
 - Validation loss decreased significantly from **2171.57** to **60.19**, showing model convergence.

Key Takeaways:

- The final accuracy (**60%**) is moderate, suggesting possible model adjustments:
 - **Data Preprocessing:** Check data normalization and scaling.
 - **Model Tuning:** Consider changing the learning rate or network architecture.
-

2. Machine Learning Models Evaluation

Several models were tested using GridSearchCV to find the best hyperparameters. The results include:

Model Accuracies:

- Logistic Regression: 60%
- Decision Tree: 80% (Best Performer)
- Random Forest: 80%
- SVM: 60%
- KNN: 40%
- Naive Bayes: 40%

Best Hyperparameters:

- Logistic Regression: $C=1$
 - Decision Tree: `max_depth=7`
 - Random Forest: `max_depth=10, n_estimators=50`
 - SVM: `C=1, kernel='linear'`
 - KNN: `n_neighbors=3`
-

3. Feature Selection Results

Feature selection was performed using two methods:

1. **SelectKBest (Chi-Square):** Selected features [0, 1, 3, 4, 5]
 2. **RFE (Recursive Feature Elimination):** Selected features [0, 1, 3, 4, 5]
- Re-evaluating models using selected features yielded mixed results:
 - **Logistic Regression:** 40% accuracy for both methods.
 - **Decision Tree & Random Forest:** 60% and 40% accuracy, respectively.
 - **SVM:** 60% accuracy using both methods.
-

- The Decision Tree and Random Forest models performed best. Consider further tuning these models.
2. **Feature Engineering:**
 - Use more advanced feature selection methods, such as PCA or Lasso regularization, to improve model performance.
 3. **Data Quality Check:**

- Ensure data is well-preprocessed, normalized, and free of outliers to enhance prediction accuracy.

Discussion:

1. Neural Network Results:

- **Initial Performance:** The neural network started with very high loss values and zero accuracy, which is typical for deep learning models at the beginning of training. This suggests that the model had difficulty in capturing patterns from the data during the initial epochs.
- **Improvement Over Time:** As training progressed, the model's accuracy gradually increased, reaching 60%. The drop in validation loss from 2171.57 to 60.19 demonstrates that the model was learning effectively and converging toward an optimal solution. However, the final accuracy of 60% is moderate, indicating that while some learning occurred, there is room for improvement. This could be due to factors like insufficient training data, improper normalization, or the need for more complex model architecture.
- **Possible Adjustments:** To enhance the model's performance, data preprocessing should be examined closely, especially with respect to normalization and scaling of features. Additionally, the network architecture could be adjusted, such as experimenting with more layers, different activation functions, or more advanced techniques like batch normalization to help reduce overfitting.

2. Machine Learning Models Evaluation:

- **Best Performers:** The Decision Tree and Random Forest models both achieved an accuracy of 80%, making them the best performers in this evaluation. This suggests that these models were well-suited to the data and could capture the underlying patterns better than other models.
- **Other Models:** Logistic Regression and SVM both reached 60% accuracy, showing moderate performance. KNN and Naive Bayes models underperformed with 40% accuracy, indicating that these models might not have been suitable for the given dataset, possibly due to issues like feature scaling or data distribution.
- **Hyperparameter Optimization:** The use of GridSearchCV for hyperparameter tuning helped identify the best hyperparameters for each model. For example, the Decision Tree performed best with a maximum depth of 7, while Random Forest performed well with a maximum depth of 10 and 50 estimators. These hyperparameters allowed the models to effectively fit the data without overfitting.

3. Feature Selection Results:

- **Selected Features:** Both SelectKBest (Chi-Square) and Recursive Feature Elimination (RFE) methods selected the same set of features, which suggests that these features are indeed the most informative for the model's predictions. However, the impact of feature selection on model performance was mixed.

- **Impact on Model Performance:** After feature selection, the Logistic Regression model still showed low performance with 40% accuracy. Decision Tree and Random Forest models showed a slight improvement with 60% and 40% accuracy, respectively. This suggests that some features may still be irrelevant or redundant, or that more advanced feature engineering techniques could be needed.

Code used Sample:

```
# إضافة طبقات مخفية
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.3))

# طبقة الخرج
model.add(Dense(y_train.shape[1], activation='softmax'))

# مع معدل تعلم مناسب Adam استخدام المحسن
optimizer = Adam(learning_rate=0.001)

# تجميع النموذج مع دالة الخسارة المناسبة
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])

# تدريب النموذج
history = model.fit(X_train, y_train, epochs=50, batch_size=32, validation_data=(X_test, y_test))

# تقييم النموذج على مجموعة الاختبار
loss, accuracy = model.evaluate(X_test, y_test)

print(f"Test accuracy: {accuracy}")

# التنبؤ بالفئات باستخدام النموذج
y_pred = model.predict(X_test)
y_pred_classes = np.argmax(y_pred, axis=1) # تحويل التنبؤات إلى فئات

# تحويل الفئات الحقيقية إلى فئات
y_true = np.argmax(y_test, axis=1)

from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.optimizers import Adam
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
from keras.utils import to_categorical
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Gas_Stations.csv')
X = df.drop('Classification', axis=1) # حذف العمود المستهدف
y = to_categorical(df['Classification']) # تحويل التصنيف إلى one-hot encoding

# تقسيم البيانات إلى مجموعة تدريب واختبار
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# بناء النموذج
model = Sequential()

# إضافة الطبقة الأولى (مدخلات)
model.add(Dense(128, input_dim=X_train.shape[1], activation='relu'))

# إضافة Dropout لتقليل overfitting
model.add(Dropout(0.3))
```

```
# حساب مصفوفة التشويش
cm = confusion_matrix(y_true, y_pred_classes)

# عرض الخريطة الحرارية
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=np.arange(cm.shape[1]), yticklabels=np.arange(cm.shape[1]))
plt.xlabel('Predicted')
plt.ylabel('True')
plt.title('Confusion Matrix Heatmap')
plt.show()
```

In conclusion, the Decision Tree and Random Forest models performed the best, but there is still potential for improvement in all models. Fine-tuning, more advanced feature selection, and ensuring high data quality are key areas to focus on to achieve better predictive performance.

This platform was developed using artificial intelligence and machine learning models to predict optimal locations, presenting the results through interactive maps that assist in making better investment decisions. The system utilizes Python and ArcGIS maps in developing the model on the interactive platform

Screenshots of the system:

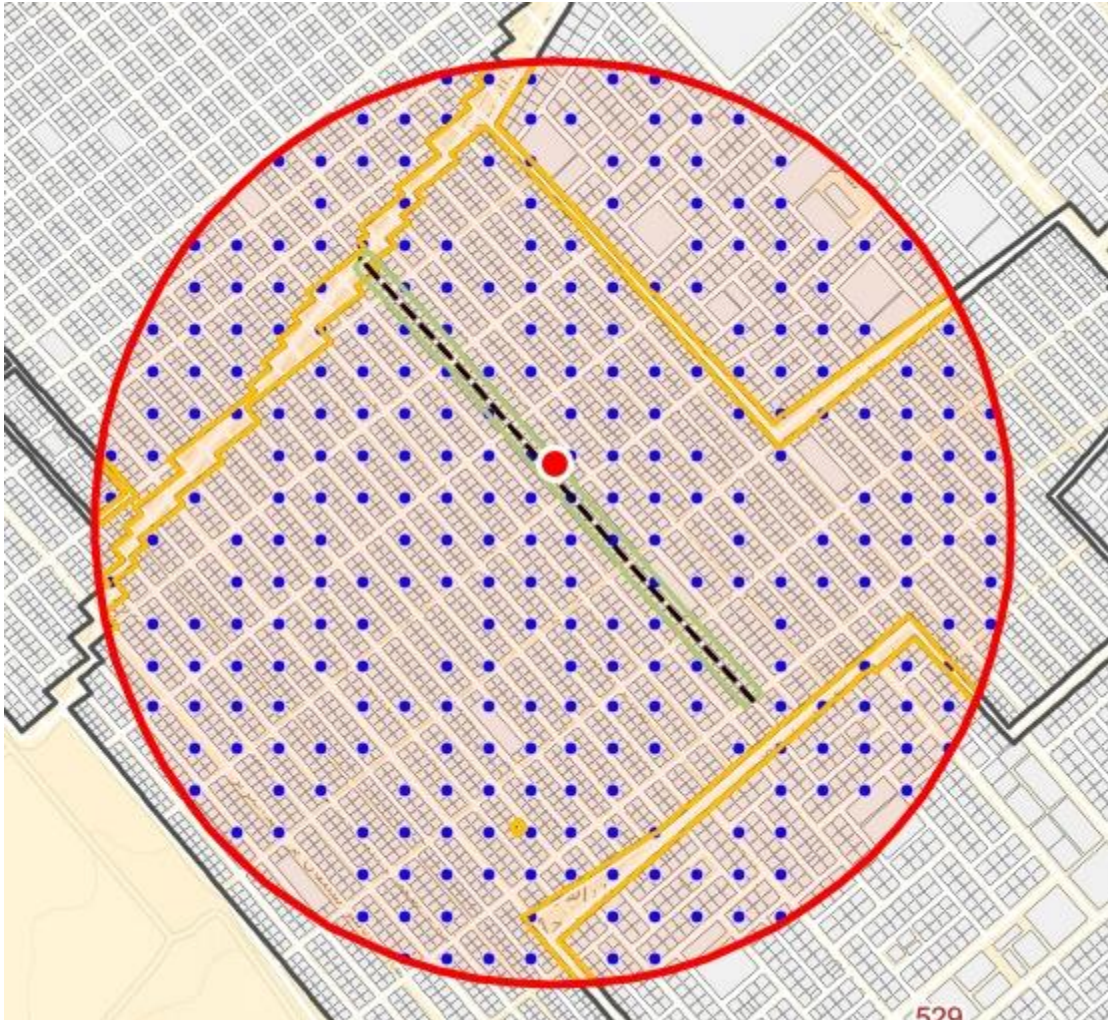


Figure (7) System screenshot of a proposed gas station location



Figure (8) Screenshots Result System

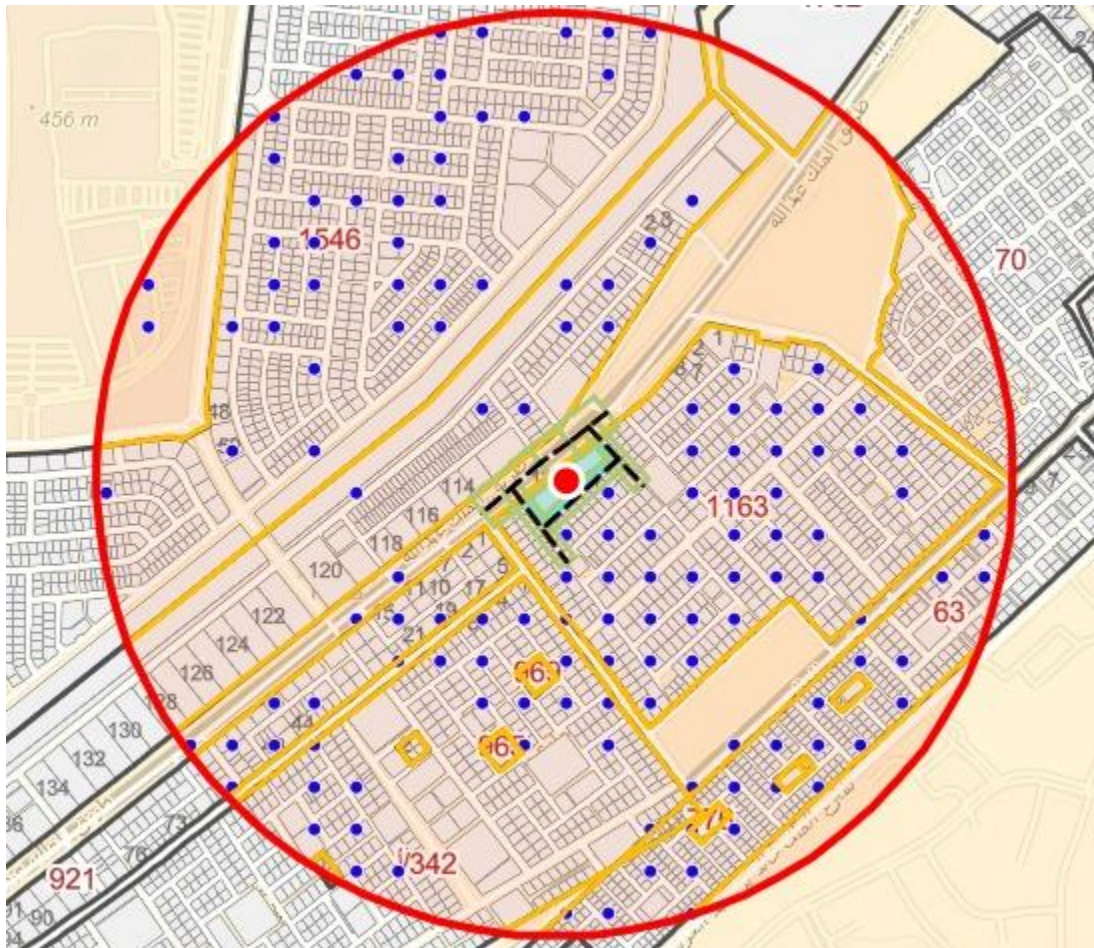


Figure (9) System screenshot of a proposed gas station location

X
النتيجة

البيانات العامة لقطة الأرض :

رقم مميز قطعة الأرض 912811718	رقم قطعه الارض Null
رقم مميز المخطط 3108	رقم المخطط 1456
الاستعمال الرئيسي الخدمات العامة	الاستعمال المحلي مسجد جامع
الاستعمال التفصيلي خدمات دينية	مساحة الارض 8436.45 متر مربع

تصنيف الموقع :

موقع متميز
موقع متوسط التميز
موقع منخفض التميز

البيانات التحليلية :

مساحة النطاق	3.14 كيلومتر مربع
عدد المحطات ضمن نطاق 2 كم	0 محطة
عدد الشوارع حول القطعه	5 شوارع
أكبر عرض شارع حول الارض	60 متر
نسبة تغطية المخططات ضمن نطاق 2 كم	80.25 %
عدد السكان ضمن نطاق 2 كم	13039 نسمة
الكثافة السكانية ضمن نطاق 2 كم	4152.5 نسمة لكل كيلومتر مربع

Figure (10) Screenshots Result



Figure (11) System screenshot of a proposed gas station location

X
النتيجة

البيانات العامة لقطعة الأرض :

رقم مميز قطعة الأرض	رقم قطعة الأرض
911937426	Null
رقم مميز المخطط	رقم المخطط
3132	61
الاستعمال الرئيسي	الاستعمال المحلي
الخدمات العامة	مستشفى عام
الاستعمال التفصيلي	مساحة الأرض
خدمات صحية	33020.13 متر مربع

تصنيف الموقع :

الموقع لا يحقق شروط محطات الوقود

البيانات التحليلية :

مساحة النطاق	3.14 كيلومتر مربع
عدد المحطات ضمن نطاق 2 كم	2 محطة
عدد الشوارع حول القطعة	5 شوارع
أكبر عرض شارع حول الأرض	50 متر
نسبة تغطية المخططات ضمن نطاق 2 كم	86.31 %
عدد السكان ضمن نطاق 2 كم	23880 نسمة
الكثافة السكانية ضمن نطاق 2 كم	7605.1 نسمة لكل كيلومتر مربع

Figure (12) Screenshots Result

Chapter 7: Conclusion and Future Work

This study addressed the analysis of factors influencing the optimal location selection for fuel stations with the aim of maximizing revenue using spatial analysis techniques and predictive models. Data related to geographic, demographic, and economic factors were collected, and analytical and predictive models were applied to identify locations with the highest potential for profit generation.

The results showed that the most influential factors include population density, traffic flow, proximity to main roads, and the availability of neighboring services. The models used demonstrated acceptable efficiency in predicting expected revenue, reflecting the accuracy of the data and the integration of the model applied. Based on the analysis, a set of potential locations meeting the selection criteria was identified, and recommendations were made for investors and business owners in the fuel station sector.

Future Work: This study opens avenues for future research in several directions, including:

1. **Developing Predictive Models:** The models used can be improved by including more temporal data, such as seasons and special events that affect fuel demand.
2. **Advanced Analysis Using Artificial Intelligence:** Machine learning techniques and neural networks can be employed to improve prediction accuracy and uncover hidden patterns in the data.
3. **Geographical Expansion:** The scope of the study can be expanded to include different regions or countries with diverse geographical and demographic characteristics to generalize the results.
4. **Environmental Impact Assessment:** Sustainability criteria can be integrated to identify locations that minimize negative environmental impact.
5. **Legal and Planning Considerations:** Study local laws and urban planning policies that influence the licensing and operation of fuel stations.

In conclusion, this research contributes to enriching the scientific literature on commercial site selection and provides a practical framework that can be applied to other commercial fields requiring decisions on optimal location determination.

References

- ✓ Fuel Station Requirements Guide - Ministry of Municipalities and Housing
- ✓ Population Density – General Authority for Statistics
- ✓ Urban Planning – Riyadh Municipality
- ✓ Streets and Roads – Ministry of Transport and Logistics
- ✓ International Journal of Economic and Environmental Geography (IJEEG)