Research Proposal

Title:
Analyzing and Predicting Hybrid Vehicle Adoption in Ireland Using Data Mining Techniques.

Name: SALMAN KHAN MOHAMMED
Student ID: 3172809

Name: SUHAIL SHAIK MUHAMMAD
Student ID: 3171799

Supervisor: Tainxiang Wang
Data: 19/11/2025

1. Introduction

Brief background of the topic: Ireland is moving toward cleaner and more environmentally friendly transport. More people are choosing hybrid cars because fuel prices are rising, environmental regulations are becoming stricter, and individuals want to reduce their emissions.

The Central Statistics Office (CSO) provides a dataset called TDM01, which shows the number of vehicles registered each month from 2006 to 2025. It includes details such as fuel type (hybrid, petrol, diesel, or electric) and vehicle category.

Because this dataset is from 19 years, it is particularly useful for examining how hybrid vehicle registrations have evolved over time. Using data mining techniques, this project will analyse these trends and create models to predict future hybrid car adoption in Ireland.

Why this problem is important:

There is currently no proper study that looks at hybrid vehicle adoption in Ireland using official national data.
This project will fill that gap by:

- Understanding how hybrid vehicle registrations have changed over the last 19 years
- Studying long-term trends in hybrid usage in Ireland
- Building forecasting models to predict future hybrid vehicle adoption
- Providing useful insights that can help with environmental planning and transport policy in Ireland.

The Research context (field, current trend, problem area):

Field: This project is in the field of data mining and sustainable transportation. It combines data analysis techniques with real-world vehicle registration data to study how Ireland is shifting toward cleaner mobility.

Current Trend: Ireland is experiencing a strong move toward eco-friendly vehicles, especially hybrid cars.
Hybrid registrations have been increasing due to:

- government climate policies
- rising fuel prices
- the shift away from diesel
- public demand for cleaner vehicles

As a result, hybrid vehicles have become a major focus in Ireland's journey toward reducing emissions.

Problem Area: Hybrid vehicles are increasing in Ireland, but no study has fully examined how hybrid registrations have changed over many years.
We still do not know:

- How hybrid vehicle adoption has grown from 2006 to 2025
- When the biggest increases happened
- What the long-term pattern of hybrid usage looks like
- How many hybrid vehicles Ireland might use in the future

Main aim and objectives: The main aim of this project is to analyze and predict hybrid vehicle adoption in Ireland from 2006 to 2025 using data mining techniques. To clean and prepare the TDM01 dataset for analysis.To analyze yearly and monthly trends in hybrid vehicle registrations.To identify the major growth periods of hybrid vehicle adoption in Ireland.To build predictive models (regression and time-series) to forecast future hybrid registrations.To evaluate the accuracy of the prediction models using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). To provide insights that help understand the long-term adoption of hybrid vehicles in Ireland.

2. Literature Review

Overview of existing work: studying on hybrid and electric vehicle shows the fuel price government incentives, and environmental awareness strongly influence consumer choices. time-series analyses are often used to identify trends. while predictive models like ARIMA, linear regression, and machine learning methods forecast demands. model accuracy typically measured using RMSE and MAE.

What others have done (main methods, findings, limitations): Other researchers has used past vehicle to study hybrid and electric vehicle adoption and they try to predict future registrations using models like ARIMA, regression and machine learning. and they found that fuel prices, government incentives, and environmental concerns affect adoption. however, most studies focus on short-term data or other countries, and many models do not include all importand factors or compare enough. in our study we are basically using irish data to predict, so its nuclear which approach works best for Ireland.

Identify research gap and justify the need for this research: Most studies on Hybrid vehicle adoption focus on short-term data or other countries with few using ireland long-term data TDM01 dataset. Model comparions and accuracy assessments using RMSE and MAE are also limites. This reasearch filles these gaps by analysing long-term irish data and eveluating prediction models to identify the most reliable approach.

3. Research Questions / Hypotheses

1. How have hybrid vehicle registrations in Ireland changed over the past 19 years?

2. Which predictive model (ARIMA, linear regression, decision trees, or machine learning) provides the most accurate forecasts for hybrid car adoption in Ireland?

3. How well do these models perform when evaluated using RMSE and MAE?

4. Methodology

Describe your approach (quantitative / qualitative / mixed): This study follows a quantitative approach, utilizing statistical analysis, data mining techniques, and predictive modeling on a long-term official dataset.

Data collection (what data, where, how)

Source: Central Statistics Office (CSO), dataset TDM01

Time Period: January 2006 – December 2025

Variables:

- Year and month
- Fuel type (Hybrid, Petrol, Diesel, Electric, etc.)
- Number of vehicle registrations

Data Preparation Steps:

- Filter hybrid vehicle registrations
- Convert date fields into time series format
- Handle missing values
- Aggregate monthly and yearly totals

Data analysis (algorithms, models, evaluation metrics):
This project will use the following steps:

1. Exploratory Data Analysis(EDA): In this we are using EDA for visualize monthly and yearly hybrid registration trends and compare hybrid vs non hybrid growth from the year 2006-2025 and we are going to identify major increases or declines

2. Predictive models: we are using linear regression for baseline trend models and ARIMA full form is (Auto-regressive integrated Moving Average) for time series forecasting and decision tree regressor for non-linear modelling and random forest regressor for ensemble machine learning.

3. Evaluation metrics: we are going to use models as RMSE (Root mean square error) and MAE (mean absolute error).

Python Tools and Implementation: All analysis will be conducted in **Python**, which is well-suited for data mining, modelling, and forecasting. The project will be developed in a **Jupyter Notebook**, with version control maintained through GitLab.

Python Libraries

- **pandas** – Data cleaning, merging, aggregation
- **numpy** – Numerical operations
- **matplotlib / seaborn** – Data visualization
- **scikit-learn** – Regression, decision tree, random forest models
- **statsmodels** – ARIMA and time-series forecasting
- **pmdarima** – Auto ARIMA tuning

Expected Output:

Identification of long-term hybrid adoption trends. Visual trend analysis from 2006–2025. Forecasts for the next 12–24 months. Accuracy comparison of four prediction models, Insights for sustainability planning and government policy.

5. References

- CSO Ireland. (2025). *TDM01: Vehicle Licensing Statistics*. Central Statistics Office.
- Higgins, C., Mohamed, M., & Ferguson, M. (2017). Electric and hybrid vehicle adoption trends. *Energy Policy*.
- Liu, W., Lin, Z., & Greene, D. (2020). Modelling consumer adoption of hybrid vehicles.
- Li, Y., & Chen, S. (2019). Predicting electric vehicle demand using ARIMA and machine learning.

Peer Feedback – Group 1:

Akash Reddy (3176643) and Srijan (3180106)

The topic is relevant and interesting. However, the research gap could be expanded with more Irish-specific references.

And they selected a good dataset from data.gov.ie Title

name is very good

Suggestions for Improvement

In future you can take whole European dataset for hybrid vehicles and in your documentation add some images of your analysis.

Peer Feedback – Group 2:

Sona Pulickal Shaju (3172233) and Oshin Elizabeth John (3172124)

The project idea is good and uniquely focused on Ireland's long-term hybrid vehicle adoption.
The topic fits the course very well because it uses data mining, forecasting, and real-world datasets
You could also mention why using 19 years of national data makes your work more valuable.
Add 1–2 more references about hybrid or sustainable mobility in Ireland to strengthen the background.

Peer Feedback – Group 3:
Muhammed Abdullah (3158419)

Strong problem statement with good real-world relevance.

Hybrid adoption is a major topic in sustainability and transportation.

The proposal is clear, but the literature review is short.

And u and can add more questions.

So now we are working on to improve ours peer feedback from the three groups they provide to us on the our Proposal.

Research Proposal Final Draft

Title:
Analyzing and Predicting Hybrid Vehicle Adoption in Ireland Using Data Mining Techniques.

Name: SALMAN KHAN MOHAMMED
Student ID: 3172809

Name: SUHAIL SHAIK MUHAMMAD
Student ID: 3171799

Supervisor: Tainxiang Wang
Data: 19/11/2025

1.Introduction

Brief background of the topic: Ireland is moving toward cleaner and more environmentally friendly transport. More people are choosing hybrid cars because fuel prices are rising, environmental regulations are becoming stricter, and individuals want to reduce their emissions.

The Central Statistics Office (CSO) provides a dataset called TDM01, which shows the number of vehicles registered each month from 2006 to 2025. It includes details such as fuel type (hybrid, petrol, diesel, or electric) and vehicle category.

Improved:

This 19-year dataset is especially valuable because long-term datasets allow deeper insights and more accurate forecasting compared to short-term studies.Because this dataset is from 19 years, it is particularly useful for examining how hybrid vehicle registrations have evolved over time. Using data mining techniques, this project will analyse these trends and create models to predict future hybrid car adoption in Ireland.

Why this problem is important:

There is currently no proper study that looks at hybrid vehicle adoption in Ireland using official national data.
This project will fill that gap by:

- Understanding how hybrid vehicle registrations have changed over the last 19 years
- Studying long-term trends in hybrid usage in Ireland
- Building forecasting models to predict future hybrid vehicle adoption
- Providing useful insights that can help with environmental planning and transport policy in Ireland.

The Research context (field, current trend, problem area):

Field: This project is in the field of data mining and sustainable transportation. It combines data analysis techniques with real-world vehicle registration data to study how Ireland is shifting toward cleaner mobility.

Current Trend: Ireland is experiencing a strong move toward eco-friendly vehicles, especially hybrid cars.

Hybrid registrations have been increasing due to:

- government climate policies
- rising fuel prices
- the shift away from diesel
- public demand for cleaner vehicles

As a result, hybrid vehicles have become a major focus in Ireland's journey toward reducing emissions.

Improved:

There is limited research analyzing Irish hybrid adoption trends over long periods, and no study has compared multiple predictive models using official CSO data.

Problem Area: Hybrid vehicles are increasing in Ireland, but no study has fully examined how hybrid registrations have changed over many years.

We still do not know:

- How hybrid vehicle adoption has grown from 2006 to 2025
- When the biggest increases happened
- What the long-term pattern of hybrid usage looks like
- How many hybrid vehicles Ireland might use in the future.

Main aim and objectives: The main aim of this project is to analyze and predict hybrid vehicle adoption in Ireland from 2006 to 2025 using data mining techniques. To clean and prepare the TDM01 dataset for analysis.To analyze yearly and monthly trends in hybrid vehicle registrations.To identify the major growth periods of hybrid vehicle adoption in Ireland.To build predictive models (regression and time-series) to forecast future hybrid registrations.To evaluate the accuracy of the prediction models using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). To provide insights that help understand the long-term adoption of hybrid vehicles in Ireland.

2.Literature Review:

Overview of existing work: studying on hybrid and electric vehicle shows the fuel price government incentives, and environmental awareness strongly influence consumer choices. time-series analyses are often used to identify trends. while predictive models like ARIMA, linear regression, and machine learning methods forecast demands. model accuracy typically measured using RMSE and MAE.

What others have done (main methods, findings, limitations): Other researchers has used past vehicle to study hybrid and electric vehicle adoption and they try to predict future registrations using models like ARIMA, regression and machine learning. and they found that fuel prices, government incentives, and environmental concerns affect adoption. however, most studies focus on short-term data or other countries, and many models do not include all importand factors or compare enough. in our study we are basically using irish data to predict, so its nuclear which approach works best for Ireland.

Identify research gap and justify the need for this research: Most studies on Hybrid vehicle adoption focus on short-term data or other countries with few using ireland long-term data TDM01 dataset. Model comparions and accuracy assessments using RMSE and MAE are also limites. This reasearch filles these gaps by analysing long-term irish data and eveluating prediction models to identify the most reliable approach.

Expanded:

Previous studies on hybrid and electric vehicle adoption demonstrate that government incentives, fuel costs, and environmental awareness significantly influence consumer decisions. Time-series methods like ARIMA and machine learning techniques such as random forest are widely used for forecasting vehicle demand.

Improved:

However, most existing research focuses on short-term datasets or other countries. They often lack long-term national-level datasets and do not include sufficient model comparisons.

Research Gap:

Improved:

- Very few studies use long-term Irish data (19 years).
- Limited comparison of forecasting models using RMSE and MAE.
- Lack of Ireland-specific analysis of hybrid trends.

This study fills these gaps by analysing long-term Irish data and comparing multiple forecasting models to determine the most accurate approach.

3.Research Questions / Hypotheses

1. How have hybrid vehicle registrations in Ireland changed over the past 19 years?

2. Which predictive model (ARIMA, linear regression, decision trees, or machine learning) provides the most accurate forecasts for hybrid car adoption in Ireland?

3. How well do these models perform when evaluated using RMSE and MAE?

4. What factors contribute most to hybrid adoption compared to other fuel types?

4.Methodology:

Describe your approach (quantitative / qualitative / mixed): This study follows a quantitative approach, utilizing statistical analysis, data mining techniques, and predictive modeling on a long-term official dataset.

Data collection (what data, where, how)

Source: Central Statistics Office (CSO), dataset TDM01

Time Period: January 2006 – December 2025

Variables:

- Year and month
- Fuel type (Hybrid, Petrol, Diesel, Electric, etc.)
- Number of vehicle registrations

Data Preparation Steps:

- Filter hybrid vehicle registrations
- Convert date fields into time series format
- Handle missing values
- Aggregate monthly and yearly totals

Data analysis (algorithms, models, evaluation metrics):
This project will use the following steps:

4. Exploratory Data Analysis(EDA): In this we are using EDA for visualize monthly and yearly hybrid registration trends and compare hybrid vs non hybrid growth from the year 2006-2025 and we are going to identify major increases or declines

5. Predictive models: we are using linear regression for baseline trend models and ARIMA full form is (Auto-regressive integrated Moving Average) for time series forecasting and decision tree regressor for non-linear modelling and random forest regressor for ensemble machine learning.

6. Evaluation metrics: we are going to use models as RMSE (Root mean square error) and MAE (mean absolute error).

Python Tools and Implementation: All analysis will be conducted in **Python**, which is well-suited for data mining, modelling, and forecasting. The project will be developed in a **Jupyter Notebook**, with version control maintained through GitLab.

Python Libraries

- **pandas** – Data cleaning, merging, aggregation

- **numpy** – Numerical operations
- **matplotlib / seaborn** – Data visualization
- **scikit-learn** – Regression, decision tree, random forest models
- **statsmodels** – ARIMA and time-series forecasting
- **pmdarima** – Auto ARIMA tuning

Expected Output:

Identification of long-term hybrid adoption trends. Visual trend analysis from 2006–2025. Forecasts for the next 12–24 months. Accuracy comparison of four prediction models, Insights for sustainability planning and government policy.

5.References

- CSO Ireland. (2025). *TDM01: Vehicle Licensing Statistics*. Central Statistics Office.
- Higgins, C., Mohamed, M., & Ferguson, M. (2017). Electric and hybrid vehicle adoption trends. *Energy Policy*.
- Liu, W., Lin, Z., & Greene, D. (2020). Modelling consumer adoption of hybrid vehicles.
- Li, Y., & Chen, S. (2019). Predicting electric vehicle demand using ARIMA and machine learning.
- **Dudeja, A., & Gupta, R. (2021). Forecasting hybrid vehicle sales using machine learning methods. *Journal of Cleaner Production*.**