Keywords: Decarbonization, Machine Learning, Time-Series Forecasting, Renewable Energy, Carbon Emissions, Road Transport.

Abstract: This article explores assessing the impact of the decarbonisation of the transport sector using an evidence-based approach incorporating data analysis and advanced machine learning (ML) modelling. We investigate the radical behavioural and societal changes needed for the decarbonisation of the transport sector in Ireland. We perform a study through our system DECARbonisation in Road Transport (DECART), a suite of statistical and time series ML models for facilitating policy making, monitoring and advising governments, companies and organisations in the transport sector. Based on data analysis and through scenario-modelling approaches, we present alternatives to policy and decision makers to achieve goals in mitigation of carbon emissions in road transport. The models depict how changes in mobility patterns in road transport affect CO2 emissions. Through insights obtained from the models, we infer that renewable energy in Ireland has the potential for meeting the growing electricity needs of electric vehicles. Experimentation is conducted on real-world datasets such as traffic, motor registrations, and data from renewable sources such as wind farms, for building efficient ML models. The models are validated in terms of accuracy, based on their potential to capture hidden insights from real-world events and domain knowledge.

1 INTRODUCTION

Globally in 2020, we saw a dramatic fall in carbon emissions, due to Covid-19 pandemic which resulted in periodic lockdowns. This was also true for Ireland where we saw around 15% decline in carbon emissions. However, with the resumption of normal activities and similar mobility patterns like those of pre-pandemic times, we are once again faced with the challenge of carbon emissions in transportation.

The statistics of greenhouse gases in Ireland (Summary by Gas | Environmental Protection Agency, n.d.) states that the major contributors are Carbon Dioxide (CO2) and Methane (CH4). CO2 mainly comes from the combustion of fossil fuels and emissions from the transport sector, contributing 40% of total fuel powered emissions in Ireland, while CH4 comes from agricultural livestock.

![Figure 1: Carbon emissions according to Vehicle types.](https://orcid.org/0000-0002-8182-2465)

![Figure 1](https://orcid.org/0000-0002-9699-522X)

![Figure 1](https://orcid.org/0000-0002-7002-5815)
transportation in Ireland. We observe that private cars lead the graph by emitting the maximum Carbon Dioxide which is then followed by Heavy Goods Vehicles and Light Goods Vehicles. Therefore, reducing emissions from transport sector is of the significant interest.

1.1 Contributions

This research paper examines how the government in Ireland can provide pathways towards a greener Ireland by decarbonizing 80% of the country by 2050. In this paper we:

1. Develop DECarbonisation in Road Transport (DECART), a Machine Learning (ML) framework leveraging real-world datasets such as traffic, motor-registrations, renewables, etc. to obtain critical and hidden insights in the decarbonisation of transport sector;
2. The DECART system will be used to monitor and plan progress in reducing carbon emissions caused by transport through predicting growth of electric vehicles;
3. The DECART system will be able to predict the potential of renewable energy to match up with the electricity requirements for the growing footprint of electric vehicles.

The structure of the research work is as follows: In the background section, we explore some of the recent works in decarbonisation using ML. In the methodology section, we discuss the architecture of the DECART system and, then we discuss the evidence-based insights based on the models in Results section. Finally, we conclude the article with key insights and recommendations.

2 BACKGROUND

There has been growing interest in the field of transportation among researchers. Some have focused on reducing congestion control through cost pricing models (Kshirsagar et al., 2021) while others have also explored alternative ways in which renewable energy can be harnessed through road traffic (Kshirsagar et al., 2021). Some works also investigated optimizing the usage of fossil fuel in the transport sector (Gota et al., n.d.; The Challenge of Decarbonizing Heavy Transport Samantha Gross Executive Summary, 2020). Researchers have harnessed the power of ML to develop forecasting models to gain valuable insights and understandings on the impact of carbon emissions in road transport (de Blas et al., 2020; Fu et al., 2019; Nouni et al., 2021).

In Ireland, this optimization was facilitated with the emergence of the carbon tax levied on fuel-based vehicles based to the age of vehicle and amount of carbon emitted by the vehicle. For every litre of diesel in addition to its cost of €1.70/litre, the carbon tax adds to 10.5 cents and the combination of Value Added Tax and excise duty adds another 80 cents to a litre of petrol (What Is the Carbon Tax? | Bonkers.Ie, n.d.).

Recently, certain initiatives from the government of Ireland have motivated people towards buying electric vehicles (EVs). Some of the intangible benefits; however, can be observed immediately with the purchase of EVs are:

- The cost of the cheapest EV can be anything between 15% to 50% higher than the conventional fuel-based passenger cars. But certain government grants make an attractive comparison in prices to conventional cars;
- Reduction in Vehicle Registration Tax summing up to €5000;
- Reductions in tolls of up to 50% for electric and 25% for hybrid as compared to fuel-based vehicles;
- A minimum saving in carbon tax of average of around €100-€300 and which can extend up to €1200 depending upon the age of fuel-based cars.

Figure 2: Electricity generation from various sources.

The rising number of EVs will cause an increase in the electricity demand for sourcing the vehicles. These growing demands of EVs can be significantly met with renewable energy sources in Ireland. Similarly, one such alternative for reduction in carbon emission, which is growing rapidly in Ireland and other European countries in the use of fuel cell electric buses, which are powered by hydrogen instead of electricity as a means of public transport. Fuel cell vehicles have more demand due to less refuelling time (10 mins), as compared to pure electric (180-240 mins). There have been 100 fuel cell electric buses active across Europe since 2018.
Phase 1 of the system helps us to understand the carbon emitting factors for fuel-based vehicles using ML models. With the insights the models, we propose some of behavioural changes in the society contributing towards reducing carbon emission.

Table 1: Dataset details.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Name</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Phase 1: Monitoring CO2 emissions from passenger cars</td>
<td>Quantitative statistics of passenger car (2017-20)</td>
<td>(Monitoring of CO2 Emissions from Passenger Cars, n.d.)</td>
</tr>
<tr>
<td>D2</td>
<td>Phase 2: Daily motor registrations</td>
<td>Number of EV registrations in Ireland (2016-21)</td>
<td>(SIMI Motorstats .n.d.)</td>
</tr>
<tr>
<td>D3</td>
<td>Phase 3: EnergyPro Dataset</td>
<td>Information about energy produced every 10 minutes period from wind farms spread across some Irish counties (2017 - 2018)</td>
<td>(Energy Pro - Specialist Windfarm Analysts &amp; Managers, n.d.)</td>
</tr>
</tbody>
</table>

3 METHODOLOGY

This section describes the proposed system architecture, datasets used, and the ML models designed for predictive analytics to gain insights from the mobility patterns in Ireland.

3.1 Architecture of DECART System and Datasets Used in Experimentation

![Image of DECART Architecture]

The DECART system comprises of a suite of regression and time-series forecasting models. This system, with the evidence-based insights obtained from the ML models, helps us to understand the present carbon footprint and plan causally for the future to address the challenge of mitigating carbon emissions in the transport sector. The pipeline of the DECART system, with its three phases is as illustrated in Figure 3.

Phase 2, with its time-series forecasting models predicts the growth of EVs in the near future. Phase 3 presents a case study using real-world dataset to predict the potential of Ireland to source the growing number of EVs in the future. The datasets used in experimentation and the overview of their usage in individual phases are discussed in Table 1.

3.2 ML Models Used in Experimentation

This work has made use of regression and time series forecasting models for predictive analytics.

3.2.1 Regression Models

Regression analysis is the process of mathematical modeling to deploy a relation between dependent and independent variables. We have used multivariate linear regression and random forest regression in our experimentations.

**Multivariate Linear Regression:** A mathematical model that deploys a linear relationship between a dependent variable, \( Y \) and multiple independent variables \( X \{X_1, X_2, ..., X_n\} \) (Schneider et al., 2010). The linear relation between these variables is described as in equation (1). The multivariate regression model is used in Phase 1 of the DECART.
system with dataset D1. The model predicts CO2 emission from the registration year, engine capacity and manufacture type of the fuel-based vehicle.

\[ Y = a + b_1 X_1 + b_2 X_2 + 1 + b_n X_n \]  

(1)

where,

- \( Y \): dependent variable
- \( X_i \): independent variables
- \( a \): constant (y – intercept)
- \( b_i \): regression coefficient of the variable \( X_i \)

**Random Forest Regressor:** Random Forest regression is the process of ensembling the predictions from sub-trees, sampled independently from the datasets to get accurate results. These regression models are more robust to noise, decreasing the generalization errors (Breiman, 2001). The random forest regressor, too, is used in Phase 1 of the system and gives better accuracy over multivariate regression model.

### 3.2.2 Time-series Models

The time series models are used in Phase 2 and 3 of the system for forecasting the growth of electric vehicles and estimating amount of energy generated with wind plants.

A time series model is a mathematical model of an event with a set of vectors \( x(t) \), where \( t = 0, 1, 2, \ldots, n \) measured over a period of time to predict the patterns of the event in future. The process of fitting a model to the data points of given time series is termed as time series analysis (Hornik et al., 1989). Over the Phases 2 and 3 in DECART system, we have used following three models for forecasting:

**Autoregressive (AR) Time Series Model:** In AR Model, a regression model is used to predict the values at given time \( t \), from the previous data points. The equation for AR model is as explained in equation (2) (Box, 1989). We have used AR model in forecasting the energy generation from wind plant with dataset D3 in Phase 3 of the system.

\[ x_t = b_0 + b_1 x_{t-1} + \epsilon_t \]  

(2)

where:

- \( x_t \): value of event at time \( t \)
- \( b_0 \): Intercept at y – axis
- \( b_1 \): slope coefficient
- \( x_{t-1} \): value of event at time \( t - 1 \)
- \( \epsilon_t \): error term
- \( t \): time

**Autoregressive Integrated Moving Average (ARIMA):** An ARIMA model used the previous data points and previous error points in a regression model to predict the values at given time \( t \). The mathematical representation of ARIMA models is as in equation (3) (John & Cochrane, 1997). ARIMA model was used for predicting the energy generation from wind plants in Phase 3 which performed better than the AR model.

\[ \Phi(L)(1-L)^d y_t = \Theta(L) \epsilon_t, \text{i.e.} \quad \left( 1 - \sum_{i=1}^{p} \phi_i L^i \right) (1-L)^d y_t = \left( 1 + \sum_{i=1}^{q} \theta_i L^i \right) \epsilon_t \]  

(3)

where:

- \( p \): order of autoregressive part
- \( d \): order of integrated part
- \( q \): order of moving average part

**Seasonal ARIMA (SARIMA):** AR and ARIMA models are used with stationary data, where the trends in the data does not vary with time. However, for forecasting the seasonal trends in the data, SARIMA models are used (Agwata Nyamato et al., 2020). The data is first converted to stationary format with pre-processing and then the future values are forecasted using the following equation (4):

\[ \Phi_p(L^d) \Phi(L)(1-L)^d y_t = \Theta_q L^d \Theta(L) \epsilon_t, \text{i.e.} \quad \Phi_p(L^d) \Phi(L) z_t = \Theta_q L^d \Theta(L) \epsilon_t \]  

(4)

where:

- \( z_t \): seasonally differenced series

The dataset D2 for predicting the rate of growth in electric vehicles in Ireland follows a seasonal trend. Hence, SARIMA model was designed to fit this data.

**ARMAX:** In some of the instances of time series models, the models are not only affected by past values of the series, but also by the external factors (Tang et al., 2000). e.g. Effect of change in wind speed on the generation of wind energy. In such cases, ARMAX models are used. We have used ARMAX models for prediction of wind energy in Phase 3 of DECART system.

**Artificial Neural Networks (ANN):** ANNs are widely used along with AR, ARIMA and SARIMA models for time series forecasting due to their unique characteristics of not defining any assumption of the data before training. Due to this significant property, ANNs can be used for large dataset with greater accuracy and universal approximations. Single hidden layer neural networks are generally used for time series forecasting (Khashei & Bijari, 2010). We have used ANNs in Phase 2 and 3 of DECART system.
3.2.3 Performance Metrics

The models were validated using performance metrics such as RMSE and R2 score. **Root Mean Squared Error (RMSE):** The Square root of average squared deviation of forecasted values is termed as RMSE. This measure gives an overview of error during forecasting (Chicco et al., 2021). The mathematical representation of RMSE is as given is equation (5).

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  (5)

where:
\( \hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots, \hat{y}_n \); predicted values
\( y_1, y_2, y_3, \ldots, y_n \); observed values
n: number of observations

**R2 Score:** R2 score or the coefficient of determination is the measure of the variance of the dependent variable with respect to the independent variable. In simple words, it is the measure of how well the prediction values lie on the line of regression (Wright, S., 1921). R2 score is defined as in equation (6).

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]  (6)

where:
\( \hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots, \hat{y}_n \); predicted values
\( y_1, y_2, y_3, \ldots, y_n \); observed values
n: number of observations

4 RESULTS AND DISCUSSIONS

This section discusses several evidence-based insights gained from the data.

4.1 Machine Learning Models for Carbon Emission Estimation

Phase 1 of the system uses ML models to understand the mobility patterns in Ireland and factors contributing towards maximum carbon emission from road transport.

The increasing emission of carbon can be addressed by transitioning towards the alternative options lowering the carbon emissions. One such alternative is the use of EVs over fuel-based vehicles. Although the transition towards electric and hybrid vehicles will be a gradual process, *certain behaviours adopted in society such as recent inclusion of hydrogen fuelled public transport, and motivating society towards maximising the usage of e-bikes* can accelerate in lowering levels of carbon emissions.

In order to predict the growth of carbon tax with the age of the vehicle, we designed a ML model to estimate the amount of CO2 emitted from a passenger vehicle from the following details: {manufacturer name, year of registration, engine capacity}. Dataset D1 was used for fitting a regression model to predict the CO2 emitted. After some preliminary experiments, we found that the best performing model was random forest regressor with an R2 score of 0.86 and RMSE of 0.36490. Overtraining was avoided by dividing the dataset into training and testing sets in the ratio of 75% and 25% respectively. The ML model will predict the gradual increase in cost of maintenance which will be directly proportional to carbon tax calculated according to age of vehicle.

The dataset was cleaned and pre-processed to fill in the missing values, converting the categorical variables into numerical ones, followed by standardisation and scaling. We used K-Nearest Neighbour Imputer to fill in the missing values. The highly correlated features were extracted using correlation matrix with heat maps which led to important features for training: {year of registration, manufacturer name and engine capacity}. The amount of carbon emitted was the predictor feature.

After training the model, we predicted the carbon emissions for the top two manufacturers in Ireland, Audi and Subaru for the vehicles registered in 2017 over a period of next 3 years. For Audi vehicles, the carbon emission increased by 0.7% in one year, but increased by 3.71% and 2.56% in the following years. Similarly for Subaru, the carbon emission was neutral for one year, however, increased by 3.20% and 1.90% in the next two years as illustrated in Figure 4.

Figure 4: Multivariate regression analysis to predict CO2 emissions for two top manufacturers in Ireland – Audi and Subaru.
The levels of carbon emission for the top two manufacturers, Audi and Subaru, were predicted from the machine learning model which led to predict the cost of carbon tax of such vehicles taking into consideration the amount of CO2 emitted from such vehicles and the fact of increasing at €7.50 per tonne of CO2 per year over the decade (CO2 Emissions from Cars: Facts and Figures (Infographics) | News | European Parliament, n.d.).

4.2 Machine Learning for Predicting the Growth in Electric and Hybrid Vehicles

The current trends in buying patterns of EVs and their expected growth in future was analysed in Phase 2 of the DECART system. The time span used was daily registration of electric and hybrid vehicles from 2016 to 2021. After preliminary pre-processing over the data, it was then used to forecast the growth in electric vehicles over the next couple of years. We predicted the growth of electric vehicles with two different models, SARIMA and ANN. The statistical results of both the models proved that SARIMA model performed better over ANN. This may be due to the fact as the data was only for the last five years, as the trends in the electric vehicles can only be observed since 2016. As the data increases in future, the accuracy of the models can be increased. The ANN model was designed with simple RNN and 100 epochs with 2 dense layers. Dataset D2 was used for fitting the model divided in the ratio of 85% and 15% respectively for training and testing to avoid overtraining. The performance of ANN model was less with R2 score of -0.57 and 0.046 as compared to SARIMA models with a R2 score of 0.49 and 0.79 to forecast the growth in electric and hybrid vehicles over the next two years.

![Figure 5: SARIMA Models for growth in electric vehicles.](image)

Figure 5 illustrates the growth of EVs predicted by the models. The spikes observed at the start and in the middle of each year in Figure 5 is due to the dual-registration system followed by the Irish Government from 2013 (Has Ireland’s Dual Registration Plate Changed the Way We Buy Cars, n.d.), where motorists can purchase a brand-new car with new number series in either January or July.

This dual registration policy has a huge impact on the distribution of car sales across the year, where July sales are almost at par with January sales. This leads to an inevitable drop in car sales in the months of Nov-Dec, as people prefer buying a new series. The models were able to capture all such real-world events. We can predict from the pattern that the growth in electric and hybrid vehicles will increase by 54.41% by the end of 2023. This will lead to a high demand for electricity to charge the electric vehicles in future. To meet those demands, we can make use of the energy generated from renewable sources.

4.3 Machine Learning Models for Predicting Energy from Wind Farms

The Phase 3 of DECART system calculates the potential of wind energy as a renewable source to cope up with the rising demands in electricity.

To understand the capacity of wind energy’s use in the transportation sector, the graph from Figure 6 illustrates that Ireland compares favourably with its European Counterparts in generating onshore wind energy. According to Wind Europe Annual Statistics 2019 (Wind Energy in Europe in 2019, 2018), Ireland had the highest share of wind in its electricity mix (33%), followed by Denmark, Portugal and Germany. Similarly, according to SEAI, (Renewables | Energy Statistics In Ireland | SEAI, n.d.), electricity generation in Ireland in 2020, wind energy contributed to generate 80% of clean energy.

![Figure 6: Comparative analysis of potential of on shore wind generation of Ireland against Europe.](image)
amount of energy generated from the Energy Pro (Welcome to Energy Pro - Specialist Windfarm Analysts & Managers, n.d.). The dataset was divided into 70% and 30% respectively for training and testing purposes. The experiments showed that ANN model performed better among the three. Hence, we then predicted the daily energy generated from one of the five plants given the information about the energy generated every 10 mins. The model used the data for energy generated for every 10 minutes in the last 24 hours and predicted the energy generation in the next 24 hours of the day.

After training the model, the results show that the average cumulative energy generated in 24 hours or 1440 minutes is 5809 KWh. The average consumption of electric is 65.26KWh for every 337.8 km (Compare Electric Vehicles - EV Database, n.d.), then we can safely say that a single plant can source around 89 electric cars in 24 hours.

According to Wind Energy Ireland (Facts & Stats, n.d.), there are 300 wind farms in Ireland. Considering this number, all the 300 wind farms have the potential to source the transport sector can be powered through clean energy by utilising wind farms to power 26,700 electric vehicles in a single day. Hence, wind farms can play a significant role in the decarbonisation of the transport sector.

The following Table 2 shows the statistical analysis of the models used for training the datasets and their results. The best results are highlighted in bold.

Table 2: Comparative Analysis of the ML models. Abbreviations used for Dataset D2 are Electric Vehicles (EV) and Hybrid Vehicles (HV).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model Name</th>
<th>R2 Score</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Random Forest Regressor</td>
<td>0.86</td>
<td>0.36490</td>
</tr>
<tr>
<td>D2</td>
<td>SARIMA</td>
<td>0.49 (EV)</td>
<td>0.79 (HV)</td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>-0.43 (EV)</td>
<td>0.60 (HV)</td>
</tr>
<tr>
<td>D3</td>
<td>ARIMA</td>
<td>0.916</td>
<td>57.69</td>
</tr>
<tr>
<td></td>
<td>ARMAX</td>
<td>0.67</td>
<td>52.41</td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>0.918</td>
<td>25.48</td>
</tr>
</tbody>
</table>

5 CONCLUSIONS AND FUTURE SCOPE

In this research work, we performed an in-depth analysis on the current emission levels in the transport sector in Ireland. Regression and time-series forecasting models were designed on real-world datasets. Through ML models we were able to gain critical and hidden insights on the mobility and behaviour patterns in the transportation sector. Furthermore, to achieve the ambitious plans of zero emissions in transportation, we discussed that clean energy, majorly sourced through wind farms, can be used to meet the electricity requirements of the rise in electric cars growth. Future scope of this article suggests case studies on other potential renewable energy sources with less or zero carbon emissions for achieving the goal of decarbonising transport in Ireland. Other aspects of future scope involve making the models dynamic to adapt to the changing trends in the transport sector as well as tuning and refining the hyperparameters of the machine learning models with Grammatical Evolution. Finally, we conclude that decarbonising transportation will highly impact societies and communities, leading to improved air quality, lower noise levels, less waste in form of emissions and thus hugely contributing towards better health and wellbeing. Similarly, with innovation in battery life of EVs may lead towards widespread acceptance of electric vehicles in near future.

ACKNOWLEDGEMENTS

This work is supported by Science Foundation Ireland grant #16/IA/4605.

REFERENCES


Has Ireland’s dual registration plate changed the way we buy cars? (n.d.). Retrieved January 27, 2022, from https://startmonday.ie/blog/ireland-dual-registration-plate


SIMI Motorstats - the complete online vehicle index for Ireland | Stats. (n.d.). Retrieved January 27, 2022, from https://stats.beepheep.ie/


