UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2020

Boosting Renewable Energy Technology Uptake in Ireland: A Machine Learning Approach

Sanghamitra Mukherjee, University College Dublin

WP20/27

September 2020

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Boosting Renewable Energy Technology Uptake in Ireland: A Machine Learning Approach

Sanghamitra Chattopadhyay Mukherjee [a, b, 1]

[a] UCD School of Economics, Newman building, University College Dublin, Dublin 4, Ireland.

[b] UCD Energy Institute, O'Brien Centre for Science, University College Dublin, Dublin 4, Ireland.

Abstract: This study explores the impact of socio-demographic, behavioural, and built-environment characteristics on residential renewable energy technology adoption. It provides new insights on factors influencing uptake using nearest neighbour and random forest machine learning models at a granular spatial scale. Being computationally inexpensive and having good classification performance, these models serve as useful baseline prediction tools. Data is sourced from an Irish survey of consumer perceptions of three key technologies – electric vehicles, solar photovoltaic panels, and heat pumps – and general attitudes towards sustainability, innovation, risk, and time. We demonstrate that utility bills, residence period, attitudes to sustainability, satisfaction with household heating, and perceptions of hassle have the biggest influence on current uptake. Urban areas, typically having better access to information and resources, are likely to see the biggest uptake first. Additionally, compatibility of household infrastructure, technical interest, and social approval are the most important predictors of potential uptake. These results may inform policy in other early adopter markets as well. Overall, policy makers must be cognisant of the stage of adoption their country is currently at. Accordingly, a holistic approach to tackling low adoption must include measures that not only enhance adoption capabilities via rebates and financial measures, but also support the opportunity and intent to purchase such technologies.

Keywords: Renewable energy technology adoption, consumer behaviour, machine learning, heat pumps, solar PVs, electric vehicles

JEL codes: D1, D9, O3, Q4

1. Introduction

In accordance with the IPCC's projections and targets, the European Commission's revised Renewable Energy Directive mandates that EU nations must fulfil at least 32% of their total energy needs with renewables by 2030 [1,2]. Much of this transition away from fossil fuels will involve a State-led re-

[1] *Corresponding author:* Email: sanghamitra.chattopadhyay@ucdconnect.ie, Phone: +353-1-716-2625, Postal Address: UCD School of Economics, Newman Building, University College Dublin, Dublin 4, Ireland.

organisation of energy production towards renewables and large-scale electrification of heat and transport [3]. Part of the transition will also entail the adoption of microgeneration technologies for electricity, heat, and transport such as electric vehicles (EVs)[2], heat pumps, and residential solar photovoltaic (PV) panels nudged along by market interventions to boost their uptake [3]. This work aims to deepen understanding of the latter, generating insights into consumer behaviour to facilitate and predict the use of these technologies, and outline favourable policy options to boost uptake.

As EVs, solar PVs, and heat pumps are technologies that have already been tried and tested in Europe, they are of specific interest to Irish policy makers. Norway is currently the European leader in EV uptake with the highest market share at 64% in January 2020 [4,5]. Meanwhile, Germany has the largest PV capacity installed in Europe at 49GWp by the end of 2019 [6]. In 2018, Sweden, Estonia, Finland, and Norway had the highest market penetration rates of residential heat pumps at around 25 sales per 1000 households [7]. Nonetheless, heat pumps still meet only about 5% of residential heat demand worldwide [7]. There is, however, a recent push in countries such as the United Kingdom to increase their installed heat pump capacity to as much as 19 million units by 2050 to be on target to meet their national target of net zero emissions by 2050 [8,9].

Although the EU as a unit seems well placed to meet the 2030 targets, individual countries are at different stages of renewable energy technology (RET) uptake [2,10]. Whilst Germany and Scandinavia may be considered RET giants in the European context, most other nation states are merely beginning this journey and have some way to go before they catch up [11]. Ireland too is still in the early stages with regard to the uptake of EVs, solar PVs, and heat pumps [12,13]. However, an accurate picture of adoption remains incomplete, as barriers and drivers relating to the uptake of individual technologies and their implications for electricity security are still ill understood, although some work has started to emerge in this area in the past year [14–16]. This research provides new insights for the Irish residential sector by employing appropriate machine learning models on rich primary data on consumer attitudes, socio-demographics and building characteristics, official census, and detailed geo-spatial data, to estimate the likely adoption patterns of these energy technologies at scale. The resulting outputs are expected to be useful to policymakers for rolling out incentive packages and necessary infrastructure to support the transition to innovative technologies by 2030.

Machine learning has a wide range of applications in decision-making. Several machine learning algorithms are ideally suited for prediction problems as they provide a quick, simple, and intuitive approach to identifying trends and patterns in data, which can be continuously improved with subsequent training with new data. Machine learning has been applied in the energy domain for several

[2] *Abbreviations:* EV : Electric vehicle; PV: Photovoltaic; RET: Renewable energy technology ; SAPS: Small area population statistics; CSO: Central statistics office; IG: Information gain; TP: True positive; TN: True negative; FP: False positive; FN: False negative; OOB: Out of bag; WGS: World Geodetic System

types of predictive tasks. Whilst deep learning models such as artificial neural networks, support vector machines, and genetic algorithms have been most frequently used, these are generally ‘black-box’ models and not easily interpretable [17]. We therefore chose the simpler decision tree, random forest and nearest neighbour approaches instead as they produce more interpretable solutions, such that our analysis is more meaningful for policy makers [18]. See Ghoddusi *et al.* (2019) for a detailed discussion of alternative machine learning methods and their applications in energy economics. Table 1 lists the relevant economics, management and finance literature that have influenced this work. We have grouped these into five subgroups: modelling energy demand/consumption (e.g. buildings) [19–29], predicting prices and risk management (e.g. stock markets) [30–37], forecasting problems (e.g. exchange rates) [38–42], decision support and policy analysis (e.g. credit card fraud) [43–49], and data mining and data management (e.g. social media) [50–54].

Applications	List of papers	Relevant methods used
Modelling energy demand/consumption	Walker *et al.* (2020)	Random forest
	Al-dosary *et al.* (2019)	Nearest neighbour
	Chen *et al.* (2019)	Random forest
	Chang *et al.* (2018)	Random forest
	Vialetto & Noro (2019)	Nearest neighbour
	Wang *et al.* (2018)	Random forest
	Zhang *et al.* (2018)	Random forest
	Mashhadi & Behdad (2018)	Random forest
	Ahmad *et al.* (2017)	Random forest
	Zagar *et al.* (2015)	Nearest neighbour
	Tsanas & Xifara (2012)	Random forest
Predicting prices, risk management	Durica *et al.* (2019)	Decision tree
	Dogah & Premaratne (2018)	Random forest
	Sun *et al.* (2018)	Decision tree ensemble
	Liu *et al.* (2017)	Decision tree
	Dey (2012)	Decision tree
	Fan *et al.* (2006)	Decision tree
	Fernández-Rodríguez *et al.* (2004)	Nearest neighbour
	Aparicio *et al.* (2002)	Nearest neighbour
Forecasting problems	Alvarez-Diaz & Mateu-Sbert (2011)	Nearest neighbour
	Biau & D’Elia (2009)	Random forest
	Alvarez-Diaz (2008)	Nearest neighbour
	Kumar & Thenmozhi (2006)	Random forest

	Pedersen & Satchell (2000)	Nearest neighbour
Decision support, policy analysis	Mercadier & Lardy (2019)	Random forest
	Antipov & Pokryshevskaya (2012)	Random forest
	Bozsik & Kormendi (2011)	Decision tree
	Lin *et al.* (2017)	Random forest
	Liu *et al.* (2015)	Random forest
	Sohn & Kim (2012)	Decision tree
	De Reyck *et al.* (2008)	Decision tree
Data mining, data management	Long *et al.* (2017)	Random forest
	Kusuma *et al.* (2016)	Decision tree
	Jiang *et al.* (2016)	Random forest
	Delen *et al.* (2013)	Decision tree
	Wu *et al.* (2006)	Decision tree

Table 1 Applications of relevant machine learning algorithms in economics and related domains

Finally, planning for and facilitating new consumer technologies requires a thorough understanding of the adoption and diffusion patterns of these and related technologies into a country's techno-economic system. Roger's illustration of diffusion as a social learning process is the most cited innovation diffusion framework in the literature [55]. A symmetric bell-shaped curve describes the distribution of adopters over time whilst a symmetric *S*-shaped pattern represents the cumulative number of adopters. Accumulation of knowledge results from interactions between past and potential adopters. As a result, five categories of adopters emerge at various stages of uptake – innovators, early adopters, early majority, late majority, and laggards. Figure 1 depicts this behavioural model. Importantly, whilst innovators tend to be independent decision-makers, interpersonal interactions have a much more profound impact on later adopters. Furthermore, the groups vary in their degree of resistance to change due to differing expectations, experience, and psycho-social characteristics. As adoption progresses, consumer segments tend to become relatively more price sensitive, older, and of lower social status and educational levels. Moreover, later adopters tend not to embrace novelty, have lower levels of product knowledge, and have comparatively less expertise in consumption than previous groups. Thus, motivations for uptake vary considerably depending on the adoption stage.

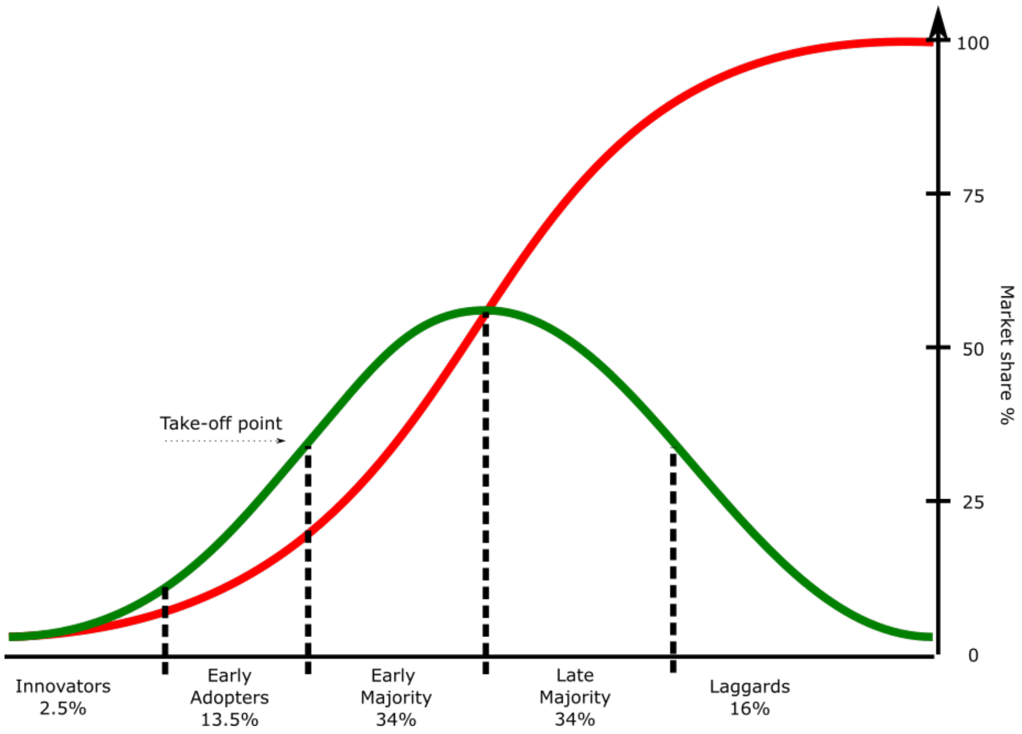


Figure 1 Technology adoption and technology diffusion curves

Source: Adapted from Everett Rogers, Diffusion of Innovations, 1962

The rest of this paper is structured as follows. Section 2 describes the data and machine learning algorithms used. Section 3 presents our RET adoption predictions for Ireland and their spatial representations. Finally, section 4 provides concluding remarks and policy implications.

2. Data and methods

This section presents the data and methods used for analysis.

2.1 Data

This work uses a combination of datasets derived from primary sources, the Irish census, and a national address database. We conducted a nationally representative online survey of Irish households to understand consumer perceptions of electric vehicles, solar PVs, and heat pumps, and general attitudes towards sustainability, innovation, risk, and time [56]. In this paper, we apply our machine learning toolkit to this rich dataset to generate the top predictors for RET adoption. Table 2 shows the representativeness of the survey data against key national statistics.

Sample attributes	% in survey	% in Ireland
Age		
18-24	11	11

	25-34	18	18
	35-44	21	21
	45-54	17	17
	55+	32	32
Gender			
	Male	49	49
	Female	51	51
Socio-economic status*			
	ABC1F50+	47	46
	C2DEF50-	53	54
Region			
	Dublin	29	28
	Leinster	26	27
	Munster	27	27
	Connaught/Ulster	18	18

Table 2 Demographic profile: Survey data

Notes: Ireland uses a system of demographic classification based on the occupation of a household's chief income earner. ABC1 collectively refers to employers, managers, and professionals, whilst C2DEF refers to non-manual, skilled manual, and semi-skilled working classes.

We used the small area population statistics (SAPS)[3] dataset associated with the 2016 national Census of Ireland to generate binary adoption/non-adoption predictions for each small area, based on the key predictors identified from the survey and the average characteristics of each small area [57]. Encompassing between 80 and 120 dwellings, the SAPS dataset is the most disaggregated level for which key socio-demographic data are available for Ireland. All socio-demographic and building characteristics that were expected to influence uptake and were available for both the census and the survey were included in the model. The remaining psycho-social and behavioural features form part of the policy frameworks that we put forward in this paper to facilitate RET uptake. Table 3 summarises the variables from the SAPS database that were used to generate predictions.

We also used geolocations from the Irish GeoDirectory database to plot the predictions on the Irish map by small area and census boundaries [58]. The GeoDirectory is a proprietary database developed by An

[3] The SAPS dataset published by the Central Statistics Office (CSO) contains data relating to sex, age, marital status, families, housing, education, commuting, and occupation, providing an in-depth picture of the socio-economic conditions for each of 18,641 Small Area communities in Ireland. The CSO has collected Census data every five years since 1926.

Post and Ordnance Survey Ireland and provides access to 2.2 million commercial and residential building addresses. Since this study is for the Irish residential sector, any purely commercial or undefined addresses were deleted, resulting in 973,316 residential addresses in our final dataset. Using the centroid coordinates for each small area, 17,747 of the 18,641 small areas were geocoded. The final dataset merging the survey, census and GeoDirectory databases was pre-processed to address several data quality issues. See Appendix A for a detailed discussion on our data cleaning and preparation process. This included feature selection to reduce the number of features as well as range normalisation and over-sampling of the survey data to avoid distortion of outputs due to measurement of input variables on different scales, and to equalise the distribution of the levels of the target variable such that the majority class does not dominate our model-based predictions. The model results presented in section 3 are for this processed dataset averaged over 100 random draws of the data.

Categorical variable	Count	% Miss.	Card.	Mode	Mode freq.	Mode %	2^{nd} mode	2^{nd} mode freq.	2^{nd}. Mode %
County	18,641	0	26	Dublin	4,882	26.19	Cork	2,186	11.73
Province	18,641	0	4	Leinster	5,144	27.60	Dublin	4,882	26.19
Occupation	18,641	0	4	Managerial, technical, or professional	11,810	63.35	Others	4,314	23.14
Employment status	18,641	0	4	At work	17,062	91.53	Unemployed, student, home-keeper, or carer	1,231	6.60
Educational level	18,641	0	5	Secondary	13,849	74.29	Third level	3,527	18.92
Gender	18,641	0	2	Female	10,798	57.93	Male	7,843	42.07
Family size	18,641	0	4	Families with up to 2 children	16,304	87.46	Families without children Parent(s) with children	2,277	12.22
Family cycle	18,641	0	4	Single- or multi-adult household	9,369	50.26	Flat or apartment	7,918	42.48
Property type	18,641	0	3	House or bungalow	16,900	90.66	Pre-1980	1,733	9.30
Building era	18,641	0	6	2001-10	8,302	44.54		6,435	34.52

(a) Categorical model inputs

Continuous Variable	*Count*	*% Miss.*	*Card.*	*Min*	*1st Qtr.*	*Mean*	*Median*	*3rd Qtr.*	*Max*	*Std. Dev.*
Number of children	18,641	0	270	0	60	90.26	87	117	552	43.74
Number of bedrooms	18,641	0	5	1	5	4.84	5	5	5	0.61

(b) Continuous model inputs before normalisation

Table 3 Summary statistics: model inputs from CSO SAPS database

2.2 Machine learning algorithms

We used two non-parametric supervised machine learning algorithms for our binary classification predictive analysis: the nearest neighbour and the random forest models. 0/1 corresponds to non-adopter/adopter classifications. All analyses were conducted using R statistical software [59]. We describe each algorithm in turn below.

2.2.1 Similarity-based learning: the nearest neighbour algorithm

The inductive bias underpinning similarity-based algorithms is that similar instances have the same target feature values. Each instance in the dataset is mapped to a point in the feature space based on the values of its descriptive features. The distance between two points in the feature space measures the *similarity* of the descriptive features of the two instances. We used the Euclidean distance to compute the distances between features. The Euclidean distance between two instances a and b in an m-dimensional feature space is defined as in *Eqn.* (1).

$$Euclidean(a, b) = \sqrt{\sum_{i=1}^{m}(a[i] - b[i])^2} \qquad Eqn. (1)$$

The steps to implement the algorithm are listed below.

Algorithm 1: Nearest Neighbour [60].

1. Define a set of training instances and a query to be classified 0/1.
2. Iterate across instances in memory and find the instance that is at the shortest distance from the query position in the feature space.
3. Make a prediction for the query equal to the value of the target feature of the nearest neighbour.

When the algorithm searches for the nearest neighbour using Euclidean distance, it partitions the feature space into a Voronoi tessellation and decides which Voronoi region the query belongs to. The Voronoi region belonging to a training instance defines the set of queries for which the prediction will be determined by that training instance. The algorithm creates a set of local models, or neighbourhoods, across the feature space, each model being defined by a subset of the training dataset. It then creates a global prediction model based on the full dataset as determined by the decision boundary[4] within the feature space.

[4] The decision boundary is the boundary between regions of the feature space in which different target levels will be predicted. It is generated by aggregating the neighbouring local models that make the same prediction.

The nearest neighbour algorithm is sensitive to noise as any errors in the labelling of training data results in errors in local models and predictions. To reduce dependency on one individual instance, we used the *k-nearest neighbour* algorithm. This model predicts the target level with the majority vote from the set of k nearest neighbours to the query q, as described in *Eqn.* (2). *Eqn.* (3) defines a distance weighted k nearest neighbour approach. The contribution of each neighbour to the prediction is a function of the inverse distance between the neighbour d and the query q, as in $\frac{1}{dist(q,d)^2}$.

$$\mathbb{M}_k(q) = \underset{l \in levels(t)}{argmax} \sum_{i=1}^{k} \delta(t_i, l) \qquad Eqn. (2)$$

$$\mathbb{M}_k(q) = \underset{l \in levels(t)}{argmax} \sum_{i=1}^{k} \frac{1}{dist(q,d)^2} \times \delta(t_i, l) \qquad Eqn. (3)$$

2.2.2 Information-based learning: the decision tree algorithm

The nearest neighbour algorithm is likely to be affected by the curse of dimensionality [61,62]. Information-based learning automatically prunes the data and builds predictive models by identifying and splitting the dataset into pure sets based on the most informative features first and may therefore be more suitable for high-dimensional data when prediction speeds are important [63]. Figure 2 illustrates a decision tree.

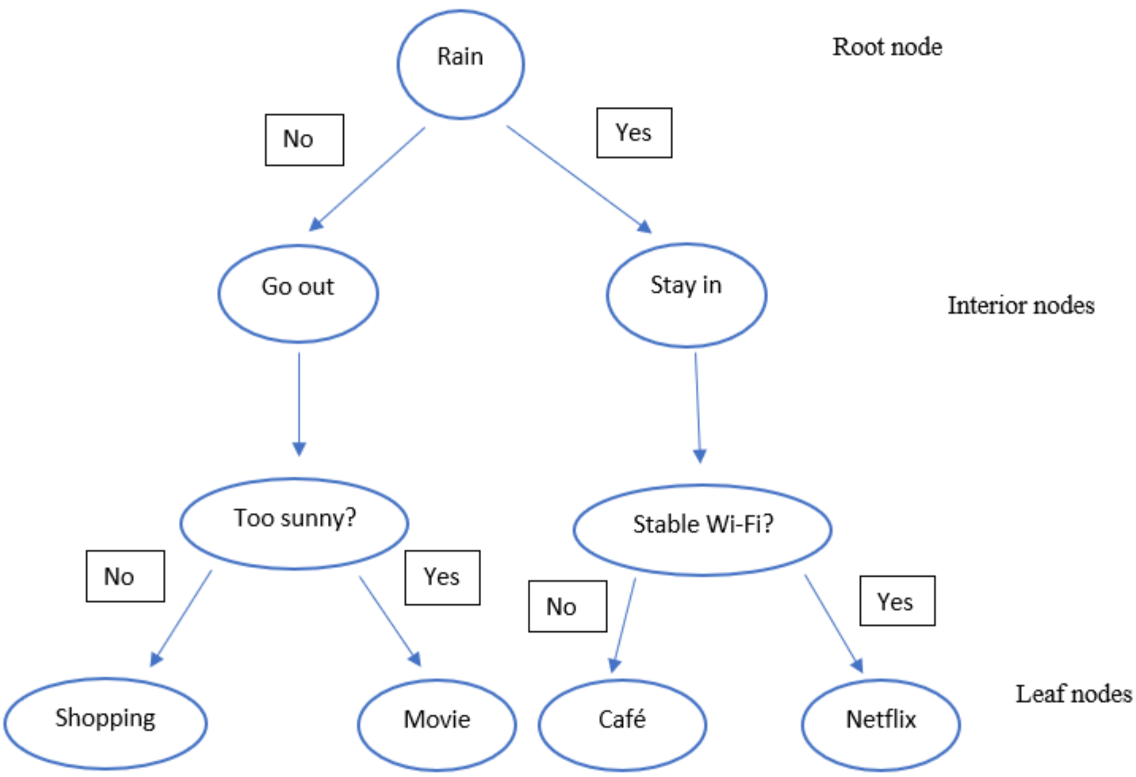


Figure 2 Sample decision tree showing root, interior, and leaf nodes

The steps to implement the algorithm are listed below.

Algorithm 2: Decision Tree [64].

1. Define a set of training instances and a query to be classified 0/1.
2. Conduct a test at each root and interior node on each of the query's descriptive features.
3. Specify a predicted classification for the query at each leaf node.

The decision tree is built in a recursive, depth-first manner, using entropy to test the purity of a set. A high outcome probability for a random selection from the set corresponds to low entropy, whilst a low probability corresponds to high entropy. We used the Shannon's entropy model presented in *Eqn.* (4), which is a weighted sum of the logs of the probabilities of each possible outcome.

$$H(t) = -\sum_{i=1}^{l}(P(t=i) \times log_2(P(t=i))) \qquad Eqn. (4)$$

The algorithm then uses information gain as a measure of the reduction in the overall entropy by testing on the target feature, calculated as in *Eqns.* (5-7).

Step 1: Calculating the entropy for a dataset with respect to the target feature:

$$H(t,\mathcal{D}) = -\sum_{l \in levels(t)} \big(P(t=l) \times log_2\big(P(t=l)\big)\big) \qquad Eqn. (5)$$

Where *levels(t)* is the set of levels in the domain of the target feature *t*, and *P(t = l)* is the probability of a randomly selected instance having the target feature level *l*.

Step 2: Computing the entropy remaining after partitioning the dataset on descriptive feature *d*.

$$rem(d,\mathcal{D}) = \sum_{l \in levels(d)} \frac{|\mathcal{D}_{d=l}|}{|\mathcal{D}|} \times H(t,\mathcal{D}_{d=l}) \qquad Eqn. (6)$$

Where $\frac{|\mathcal{D}_{d=l}|}{|\mathcal{D}|}$ denotes the weighting determined by the size of each partition, and $H(t,\mathcal{D}_{d=l})$ denotes the entropy of partition $\mathcal{D}_{d=l}$. A larger partition contributes more to the overall remaining entropy.

Step 3: Calculating the information gain (IG) from splitting the dataset $\mathcal{D}$ using the feature *d*.

$$IG(d,\mathcal{D}) = H(t,\mathcal{D}) - rem(d,\mathcal{D}) \qquad Eqn. (7)$$

The model makes a predicted classification when one of three situations occur:

1. All instances have the same classification: model returns a leaf node tree with that classification.
2. No features left to test: model returns a leaf node tree with the majority class as its classification.

3. The dataset is empty: model returns a leaf node tree with the majority class at the parent node.

We used decision trees as part of an ensemble, viz., Random Forests, first proposed in Breiman (2001) [65]. Random forests make predictions based on a set of induced decision tree models by returning the majority vote or median.

2.2.3 Model evaluation metrics

We employed a *k*-fold cross validation experiment implying that *k* separate evaluation experiments were performed. In the first evaluation experiment, the data in the first fold was used as the test set, and that in the remaining *k* - 1 folds were used as the training set. This process was repeated 10 times, i.e. *k*=10. Figure 3 illustrates the model training, testing, and evaluation process.

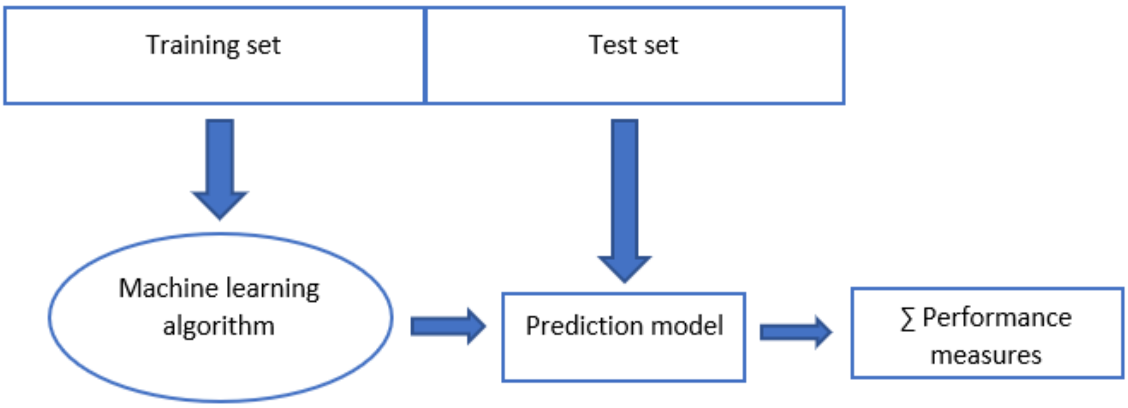


Figure 3 Workflow for predictive models

We implemented multiple performance measures as part of our evaluations to avoid erroneous judgements on the real performance of our models. *Eqns.* (8-11) list our measures of choice.

$$Classification\ accuracy = \frac{Number\ of\ correct\ predictions}{Total\ predictions}$$

$$= \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad Eqn. (8)$$

$$Recall = \frac{TP}{(TP + FN)} \qquad Eqn. (9)$$

$$Average\ class\ accuracy_{AM} = \frac{1}{|levels(t)|} \sum_{l \in levels(t)} recall_l \qquad Eqn. (10)$$

$$Average\ class\ accuracy_{HM} = \frac{1}{\frac{1}{|levels(t)|} \sum_{l \in levels(t)} \frac{1}{recall_l}} \qquad Eqn. (11)$$

Although classification accuracy is one of the most intuitive metrics, it is sensitive to class imbalance. In contrast, average class accuracy corrects for class imbalance by computing the accuracy of each class separately and averaging the results (using either arithmetic or harmonic mean) thus allowing each class to have equal weight when computing the metric. Hence, we present these first. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are components of a confusion matrix as illustrated in Figure 4. TP is an outcome where the model correctly predicts the positive class, TN is an outcome where the model correctly predicts the negative class, FP (also called a type 1 error) is an outcome where the model incorrectly predicts the positive class when it is actually negative, and FN (also called a type 2 error) is an outcome where the model incorrectly predicts the negative class when it is actually positive.

		Prediction	
		Positive	**Negative**
Target	**Positive**	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Figure 4 Structure of a confusion matrix

3. Results

This section reports our model predictions for RET uptake in Ireland and their spatial representation.

3.1 Model predictions

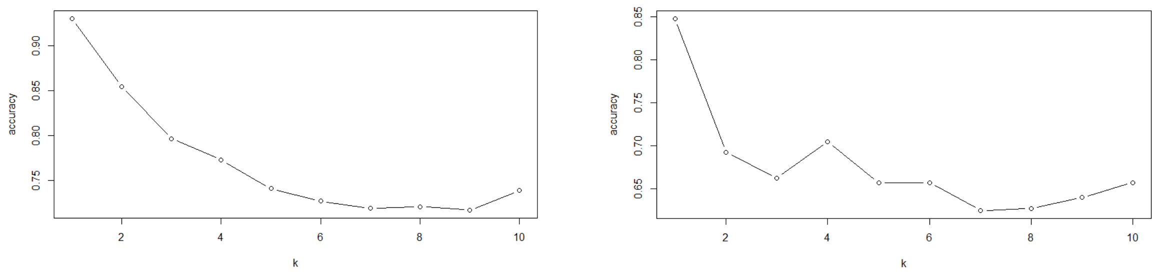
We ran the algorithms for each of our two target variables – adopters of RET and potential adopters of RET[5]. We chose not to re-run the models for individual technologies (i.e. EVs, solar PVs or heat pumps) due to low sample sizes for each subgroup. The train and test sets were split in the ratio 3:1. We did a 10-fold cross validation on the train sets and then evaluated on the test sets.

3.1.1 Similarity-based learning: the nearest neighbour algorithm

We ran the *k*-nearest neighbour algorithm using two independent variables – number of children and years of education. These were selected based on the feature selection process detailed in Appendix A.

[5] A potential adopter is defined as a survey respondent who either owns an RET (and are thus likely to purchase more RETs in the future) or demonstrated a strong interest in purchasing one (by self-proclaiming that they were likely to purchase one in the future). An adopter is defined as a survey participant who owns at least one type of RET. Both adopter and potential adopter are binary outcome variables, i.e. they take on a value of 0 for non-adoption and a value of 1 for adoption.

The optimal value of k was based on the average class accuracy for each value of k tested. Figure 5 plots this selection process. Tables 4 and 5 present the confusion matrices and model evaluation metrics for dependent variables *adopter* and *potential adopter,* respectively. At between 0.5 and 0.6, the average class accuracy metrics are not too high. However, this is unsurprising due to limitations in data availability.



(a) Adopter (optimal k = 1) (a) Potential adopter (optimal k = 1)

Figure 5 Optimal k values for the k-nearest neighbour algorithm using average class accuracy

***k* = 1**	**Predicted**		**Evaluation metrics**
Actual	*Non-adopter*	*Adopter*	Average class accuracy (arithmetic mean) = 0.608
Non-adopter	116	135	Average class accuracy (harmonic mean) = 0.573
Adopter	62	189	Classification accuracy = 0.608

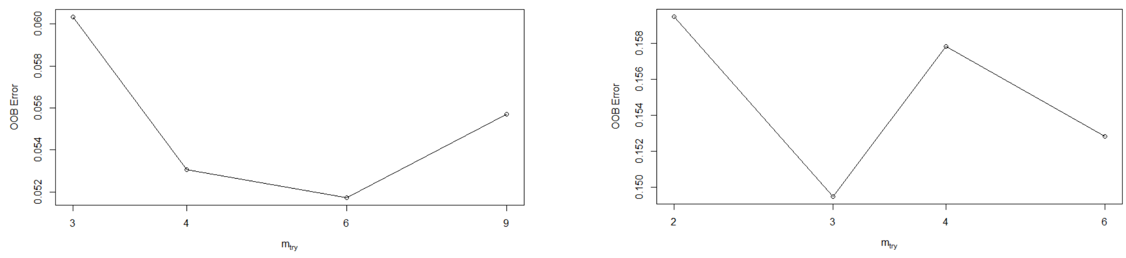
Table 4 Confusion matrix and evaluation metrics, k-nearest neighbour, *Adopter*

***k* = 1**	**Predicted**		**Evaluation metrics**
Actual	*Potential non-adopter*	*Potential adopter*	Average class accuracy (arithmetic mean) = 0.53
Potential non-adopter	120	80	Average class accuracy (harmonic mean) = 0.521
Potential adopter	108	92	Classification accuracy = 0.53

Table 5 Confusion matrix and evaluation metrics, k-nearest neighbour, *Potential adopter*

3.1.2 Information-based learning: the decision tree algorithm

We built a decision tree-based ensemble model, the random forest model, using the top 30 features (as identified in the feature importance plots for the full dataset) that had corresponding data available in the SAPS dataset[6]. The optimal number of trees generated, *ntree*, was calculated by minimising the out of bag (OOB) error rate[7]. Figure 6 illustrates the selection process for the optimal number of features in the construction of each tree, *mtry*[8], for the corresponding *ntree* values.



(a) Adopter (optimal *mtry*=6, *ntree*=100)

(b) Potential adopter (optimal *mtry*=3, ntree=900)

Figure 6 Optimal *mtry* and *ntree* values for training a random forest model for the CSO dataset

The feature importance plots for the CSO dataset are presented in Figures 7 and 8. Mean decrease in Gini impurity was used to measure information gain to rank variables in terms of importance; the absolute values are not important. Tables 6 and 7 present the relevant confusion matrices and model evaluation metrics. The models perform well, with average class accuracy figures of 89% and 83% for *adopter* and *potential adopter*, respectively[9]. Due to better model performance than the k-nearest neighbour approach, we investigate the random forest model further in Section 3.2 where we illustrate the spatial representations of the random forest model predictions. For reference, we also present the feature importance plots for the full dataset in Figures 9 and 10 as we use insights from these for our policy discussion later. The corresponding confusion matrices are presented in Tables 8 and 9. Across

[6] Refer to Table 4 for the full set of variables. The variables *children over 17* and *children under 17* were combined to form a new variable - the total number of children - due to differences in the cut-off age used to determine who a child is in the survey and CSO datasets.

[7] The OOB error rate is equal to the number of points in the training set that were misclassified divided by the total number of observations.

[8] These features were selected at random. The default value for classification is sqrt(number of features).

[9] Due to concerns around using random forests to learn imbalanced data [78], we ran the models on oversampled, undersampled, and weighted data. Undersampling and weighting produced slightly lower classification accuracies than oversampling. Although oversampling risks overfitting the data, undersampling is prone to losing important information in the majority class [79]. Since our dataset was small to begin with, we decided that oversampling was the better solution in our context.

all models, predictions for non-adopters were better than for adopters, implying that false negatives were minimised.

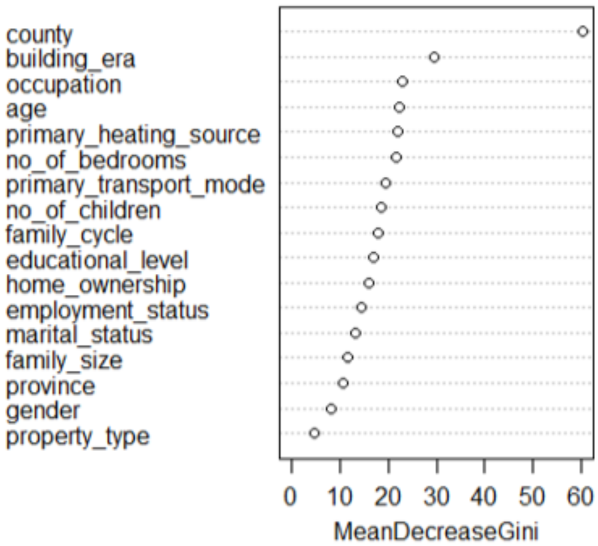


Figure 7 Feature importance as measured by a Random Forest, CSO dataset, *Adopter*

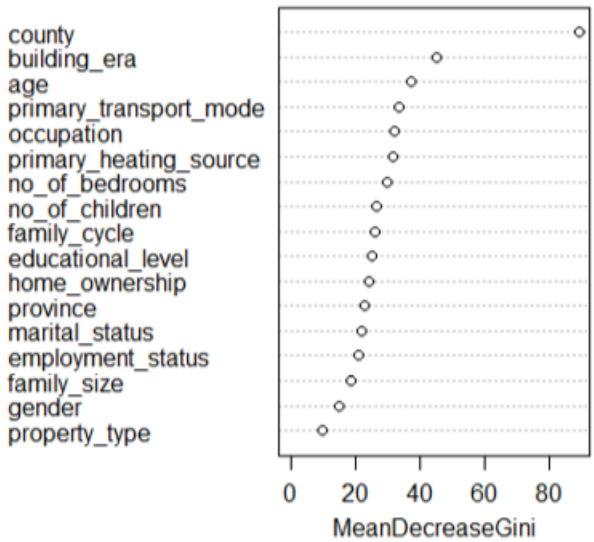


Figure 8 Feature importance as measured by a Random Forest, CSO dataset, *Potential adopter*

Random forest	**Predicted**		**Evaluation metrics**
Actual	*Non-adopter*	*Adopter*	Average class accuracy (am) = 0.892
Non-adopter	238	13	Average class accuracy (hm) = 0.889

Adopter	41	210	Classification accuracy = 0.892

Table 6 Confusion matrix and evaluation metrics for the Random Forest model, CSO dataset, *Adopter*

Random forest	**Predicted**		**Evaluation metrics**
Actual	*Potential non-adopter*	*Potential adopter*	Average class accuracy (am) = 0.838
Potential non-adopter	175	25	Average class accuracy (hm) = 0.836
Potential adopter	40	160	Classification accuracy = 0.838

Table 7 Confusion matrix and evaluation metrics for the Random Forest model, CSO dataset, *Potential adopter*

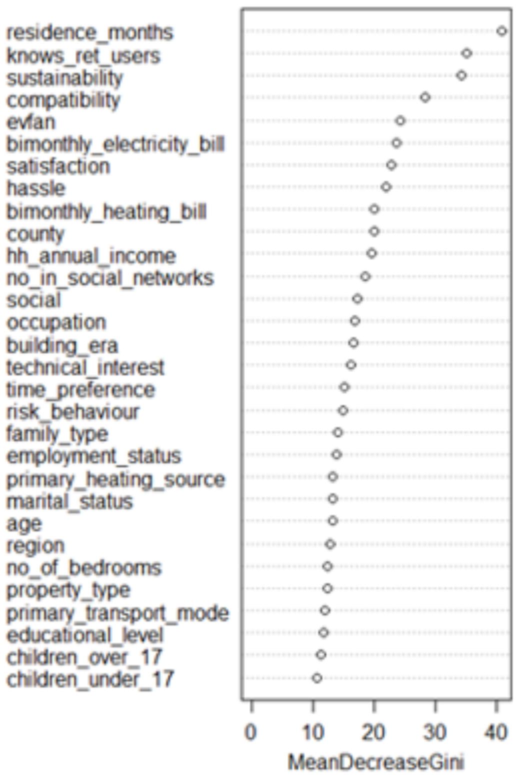


Figure 9 Feature importance as measured by a Random Forest, Full dataset, *Adopter*

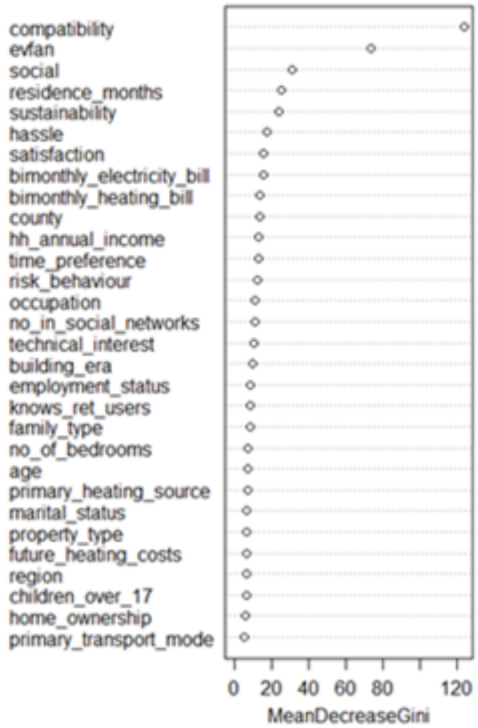


Figure 10 Feature importance as measured by a Random Forest, Full dataset, *Potential adopter*

Random forest	**Predicted**		**Evaluation metrics**
Actual	*Non-adopter*	*Adopter*	Average class accuracy (am) = 0.976
Non-adopter	251	0	Average class accuracy (hm) = 0.976
Adopter	12	239	Classification accuracy = 0.976

Table 8 Confusion matrix and evaluation metrics for the Random Forest model, Full dataset, *Adopter*

Random forest	**Predicted**		**Evaluation metrics**
Actual	*Potential non-adopter*	*Potential adopter*	Average class accuracy (am) = 0.91
Potential non-adopter	187	13	Average class accuracy (hm) = 0.909
Potential adopter	23	177	Classification accuracy = 0.91

Table 9 Confusion matrix and evaluation metrics for the Random Forest model, Full dataset, *Potential adopter*

Overall, the model for the full dataset identified residence period, knowing other RET users, sustainability, compatibility, being an EV fan, utility bills, satisfaction, the hassle factor, and county as the key predictors for current adoption, and compatibility, being an EV fan, social approval, residence period, sustainability, the hassle factor, satisfaction, utility bills, and county as the top predictors for potential adoption. Thus, behavioural factors featured prominently for both current and potential uptake. We used the socio-demographic, locational and building characteristics identified from our predictions for the full dataset for our mapping exercise that follows.

3.2 Spatial patterns of RET adoption

We used Esri's cloud-based GIS mapping software, ArcGIS Online [66], to map the random forest model predictions by small area and county boundaries and corresponding point density plots for adopters and non-adopters classified by standard deviations. We used the standard World Geodetic System (WGS84) reference coordinate system for all maps [67]. Figure 11 depicts a reference map of Ireland showing the major cities and road network in Ireland. Dublin, Cork, Limerick, Galway, and Waterford are the largest and most populous cities in Ireland in that order. The road network connects the capital city, Dublin, with other major Irish towns and cities.

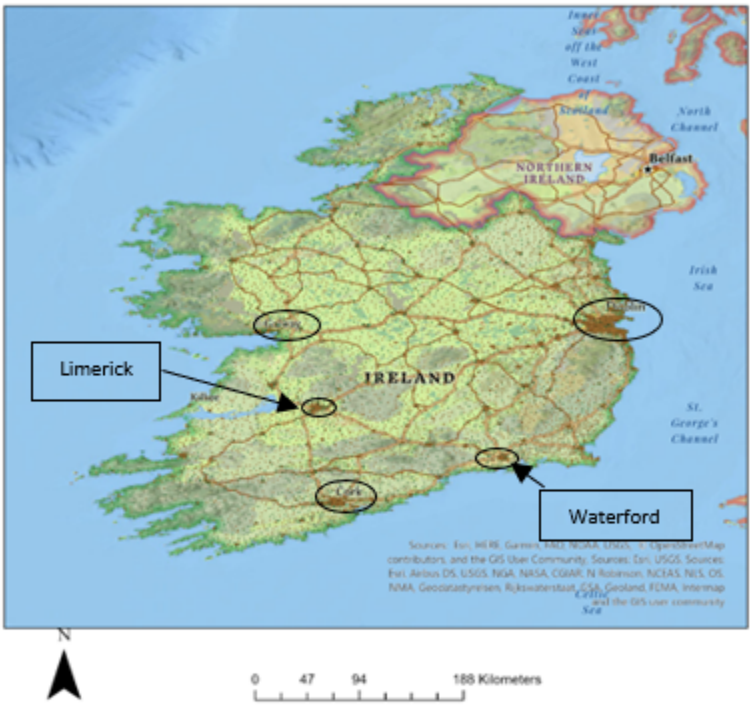
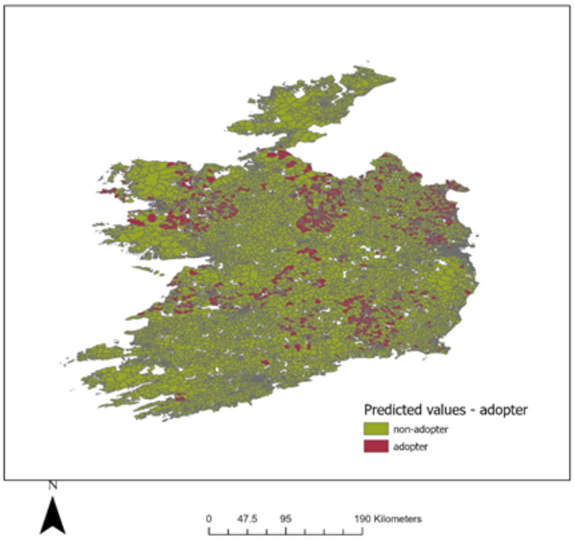
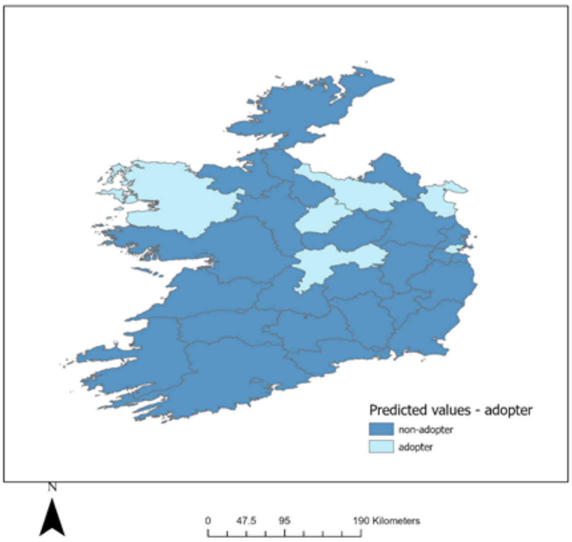


Figure 11 Map of Ireland showing small area centroids, major cities, and road network

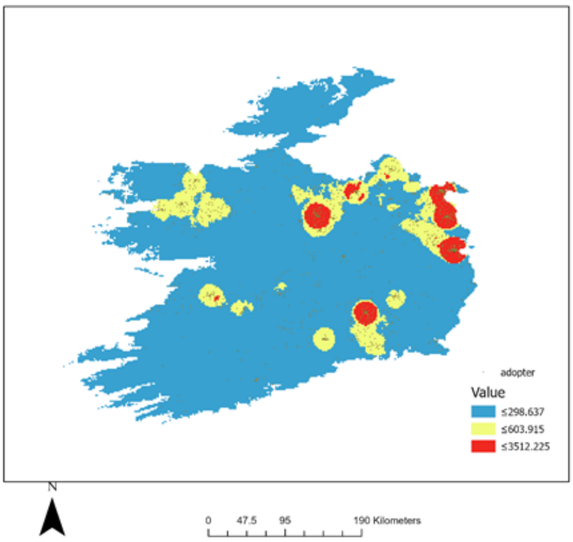
Figures 12-13 depict the spatial representations of our predicted distributions for variables *adopter* and *potential adopter,* respectively. We do not present maps of adoption by socio-demographic characteristics as these did not show any significant variation across regions. This homogeneity in socio-demographics is expected as average values were used to characterise areas. Being a reasonably small country with a small population[10], regional average values are likely to be similar.



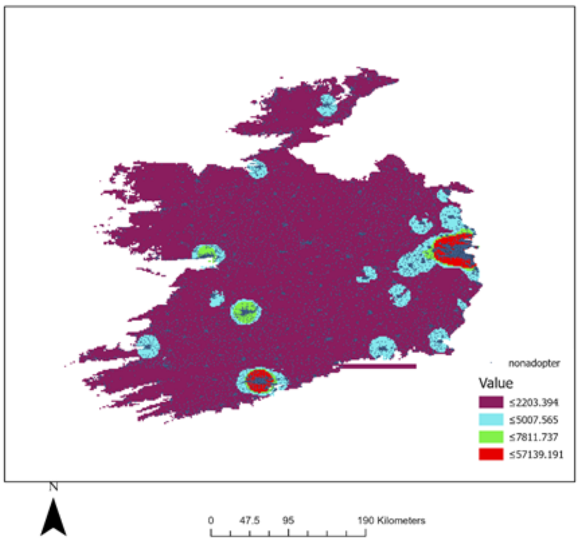
(a) Predicted values of adopter/non-adopter classifications by small area boundaries



(b) Predicted values of adopter/non-adopter classifications at county level



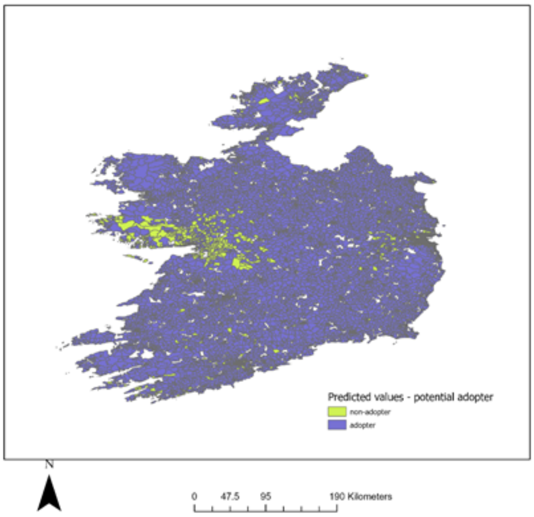
(c) Point density plots for current RET adopters by small area centroids



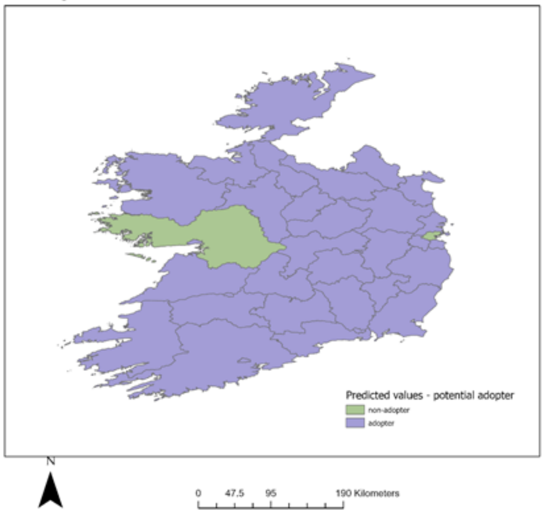
(d) Point density plots for current RET non-adopters by small area centroids

[10] The Republic of Ireland has a population of 4.9 million and a land area of 70,273 km² divided into four provinces encompassing 26 counties.

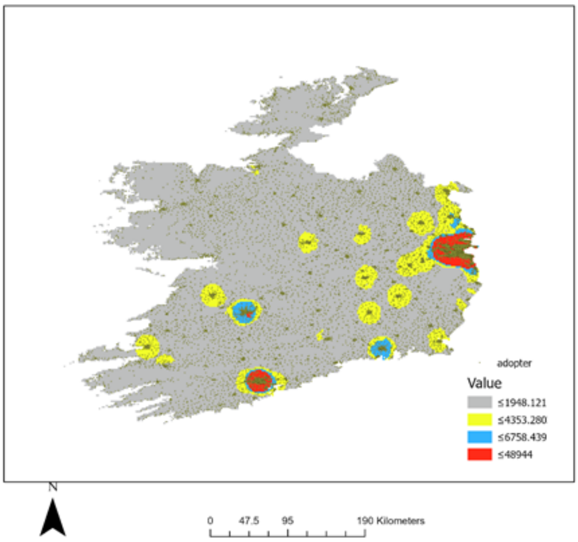
Figure 12 Adopter and non-adopter classifications, Random forest algorithm, *Adopter*



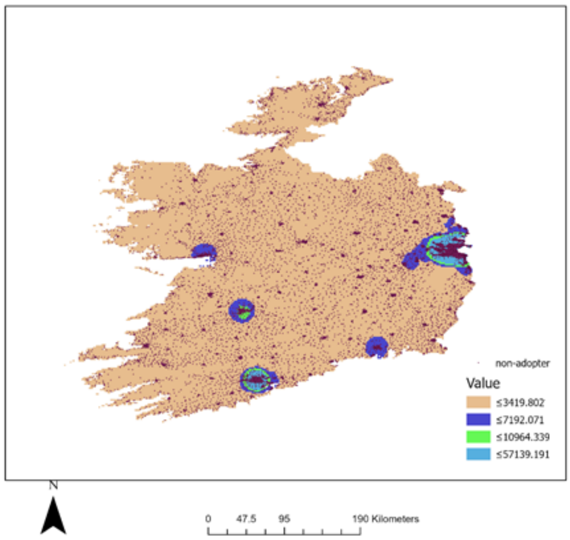
(a) Predicted values of adopter/non-adopter classifications by small area boundaries



(b) Predicted values of adopter/non-adopter classifications at county level



(c) Point density plots for potential RET adopters by small area centroids



(d) Point density plots for potential RET non-adopters by small area centroids

Figure 13 Adopter and non-adopter classifications, Random forest algorithm, *Potential adopter*

For our predicted variable *Adopter*, the model predicted that adopter regions are located primarily in the north. The density plots detected adopter clusters around Dublin and several other towns in the province of Leinster. Most non-adopters seem to cluster around the populous regions of Dublin, Cork, Limerick, and Galway, rather than Kilkenny and the north-east, regions where adopters appear to cluster. For our predicted variable *Potential adopter*, the model predicted all regions as being potential RET adopters except for areas in the west of the country surrounding County Galway and the east of the country surrounding Dublin city. Potential adopter clusters were detected around Dublin, Cork, and Limerick. Potential non-adopter clusters appeared primarily around Dublin, Cork, Limerick, Waterford, and Galway.

4. Discussion

From the key predictors identified in Section 3.1, it appears that RET adopters are not only better educated and in higher professional jobs, but are also quite satisfied with their household insulation (and thus utility bills) and heating/cooling infrastructure[11] [12] before they adopt an RET. Besides, adoption is linked to lower residence periods which may either correspond with newer houses that have the infrastructure in place to install new technologies, or point to the fact that younger generations are more willing to adopt RETs as they are typically more technically competent and/or more likely to accept a higher level of uncertainty surrounding new technology. Age, however, did not figure as a key predictor in the feature importance plots. Furthermore, families with children are more likely to currently adopt RETs. In terms of attitudinal predictors, sustainable behaviours appear to have had a negative influence in the recent stages of adoption implying that current adopters are less environmentally minded than non-adopters. Finally, current adopters display technical inclinations in terms of in-depth knowledge of EVs, for instance, and enjoy personal contact with other RET users.

These observations imply that the decision to adopt has thus far not been an outcome of dissatisfaction with previous infrastructure, high energy bills, or environmental concern, but has perhaps stemmed from an interest in new technology, an ability to pay for them, and a need based on the perceived benefits of RET for a larger family, maintaining lower utility bills and/or the positive experiences of other RET users in adopters' own social circles. Importantly, although social approval has not figured in current uptake in any meaningful way, it was predicted to play a major part in potential uptake. This may indicate that the Irish customer base is presently transitioning from the early adopter to the early majority stage, as peer-to-peer communication is much more influential in the later phases of adoption when opinion leadership diminishes markedly [55]. Overall, it is important to note that the motivations for purchasing a technology change as technologies continuously evolve and the customer base matures [68]. Therefore, any new market based instruments must be flexible and reflect the changing needs of

[11] This may be because good insulation and infrastructure are necessary *prior* to installing RETs such as heat pumps. Any additional heating and cooling benefits that these technologies provide add to the utility derived from owning them but are not influential in the decision to adopt.

[12] We confirmed the direction of effect of all variables using a logistic regression model. See Appendix B for model outputs.

the buyer [69]. Like social processes, sustainable behaviours may co-evolve with RET take-up rather than play a static role over time [70]. Moreover, given that sustainability and social influence are often intertwined, policy design may also engage thought leaders from previous rounds of uptake to encourage the diffusion of environmental behaviours in subsequent stages [68]. These insights should be useful to policymakers as more and more governments now seek to apply behavioural insights to inform policy and improve public services [71,72].

Spatial dependence is a recognised phenomenon in economic and social activity [73,74]. Spatial proximity is a key determinant of RET adoption, perhaps in the way it brings other RET users (and, thus, information and influence) and public infrastructure closer to the adopter. Our spatial analysis unearthed several insights for planners and policy makers, the implications of which are two-fold. Firstly, the maps for *Adopters* (Figure 12) illustrated the current distribution of RET adopters across the country. This includes regions that are currently adopting and those that are likely to adopt in the immediate future based on the average characteristics of current adopters. Current adopters are located primarily around Dublin, Kilkenny, and the north-east. Furthermore, clusters of adopters appear quite prominently in the province of Leinster. This makes sense as Leinster has the highest population [75] and is home to more people with high disposable incomes and third level education than other parts of Ireland [76]. Planners can compare uptake in these high uptake regions with the availability of relevant electricity grid infrastructure to plan for upgrades to strengthen capacity where needed. In contrast, most current non-adopters seem to cluster around Dublin, Cork, Limerick, and Galway. Dublin figures in both adopter and non-adopter clusters probably due to its high centred population and being the largest city in terms of size. Although Cork, Limerick, and Galway are the next biggest in size, they house far fewer people than Dublin [77] and, therefore, these cities are identifiable as non-adopter clusters possibly due to very low uptake numbers overall and sparse distribution of adopters. Policymakers have a role here to devise incentive packages to encourage uptake in these areas, such that no region is left behind due to barriers such as a lack of access to necessary resources or information. Low uptake regions may also indicate areas where idle resources could be re-structured or re-allocated.

Secondly, the maps for *Potential adopters* (Figure 13) illustrated RET adoption based on the average characteristics of potential adopters in the survey. The results showed that with supportive policies and suitable infrastructure, most regions are likely to take up RET. This is good news for Ireland's environmental commitments as it will enable the nation to move to a more sustainable future in the next decade and help meet Ireland's 2030 renewable energy targets. Overall, it is no surprise that potential adopter clusters primarily surround populous urban locations such as Dublin and Cork. Adopter densities are tied to population densities and accessibility of information and resources (likely better in urban regions), as well as connectivity with major towns, especially for EVs, which cannot be used without an extensive road network and access to public charging facilities typically located on-street, in car parks, or at filling stations alongside major roads and motorways. Further research is needed on the rural-urban divide in both current and potential RET uptake to analyse the presence of regional disparities. Either way, policies that aim to encourage uptake may need to be area-specific to be effective. We present the policy implications emanating from this research below.

Policy implications

EVs, solar PVs, and heat pumps are by this stage reasonably well established in Ireland. Owing to the importance of social processes identified in this work for future uptake, it is possible that these technologies are now in the tail end of the 'early adopter' stage, moving into the 'early majority' phase in the next five years. This knowledge is crucial as the early majority has very different needs and expectations compared with innovators and early adopters. Whilst innovators thrive on significant change and are typically wealthier and more willing to bear risk and higher levels of uncertainty, the early majority are generally not technology enthusiasts, and seek greater reliability, consistency, and predictability in the products they purchase. Thus, although most consumers would still accept a period of acclimatisation to a new technology, they would expect technologies to work efficiently and in a cost-effective manner without the need to understand the nuances of ever-changing operation and maintenance. Accordingly, familiarity with novel technologies is key, and support systems must already be in place that enhance RET buying, ownership, and after-sales experiences. Hence, an efficient policy design during this transition phase needs to enhance not only the capability of consumers in terms of their knowledge and financial ability to purchase but also their intent (and, thus, their willingness) and opportunity (and thus, the choices presented to them) to purchase and retain RET. As illustrated in Figure 14, we propose a novel three-pronged policy framework for the design of strategies to boost overall RET uptake, based on our predictive modelling and behavioural insights.

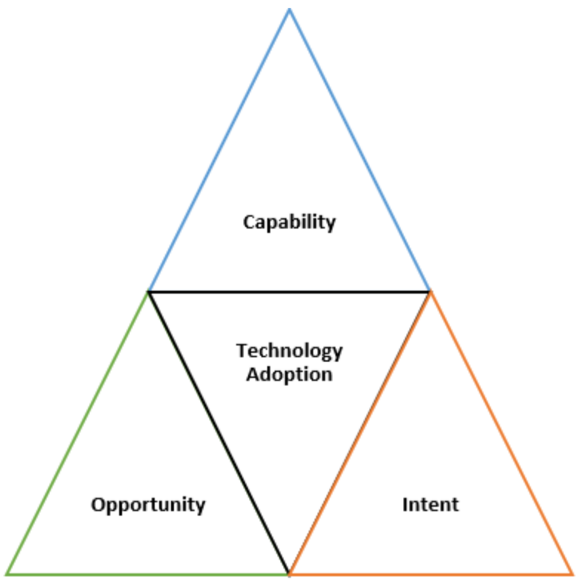


Figure 14 Intent, Capability, and Opportunity: A holistic approach to boosting RET uptake

Firstly, regarding *capability*, the early majority is typically more price sensitive than the early adopters. Thus, for technologies that are still relatively expensive, consumers would inevitably assess value-for-money primarily. Consequently, financial incentives based on consumer preferences that either subsidise RETs and/or tax fossil fuel consumption may make RETs significantly more attractive for the early majority than for any previous group of adopters. In addition, flourishing, viable and independent second-hand dealerships, such as those that sell used EVs refurbished to a recognised industry standard, would have greater appeal for this group than any others. It is favourable for EV ownership that standardised charging is becoming increasingly commonplace in Europe. However, the price to charge away from home must also remain reliable and consistent across the charging network. For solar PVs and heat pumps, homeowners have traditionally installed these technologies as part of major home renovations on the advice of building contractors. Most current adopters know very little about the make and operation of their installed units and rely largely on installers to guide their use and maintenance. Thus, when technologies are complex and/or the user base is not highly technically competent, the capability to adopt is also influenced greatly by the support of advisers. Overall, financial incentives alone are not sufficient to encourage adoption; to enhance the capability to purchase and retain RET, not only must upfront costs be made financially manageable, but the pace at which advice and assistance on installation, use and maintenance is delivered must also be quick, consistent and reliable.

Secondly, the notion of *opportunity* in purchase decisions relates to the accessibility and availability of RET-related information and infrastructure, as well as the opportunity to trial innovations before committing to purchase long-term. An extensive public charging network equipped with chargers of varying speeds and capacities within reasonable reach of a consumer would, for instance, encourage EV take-up by easing anxieties surrounding long distance commutes. It would also make EVs more accessible to communities without home charging facilities, such as for residents of compact cities and rental accommodation. Other opportunity-enhancing targets could include reducing queuing at public chargers to a minimum by introducing a tiered payment scheme, organising free technology demonstrations at regional trade fairs, extending product guarantees for new customers, and facilitating low-cost returns and swaps to promote trialability. Additionally, it is important to note that whilst most innovators and early adopters would base their purchase decisions on opportunities to actively contribute to the development of innovations, the early majority will prefer ready-made solutions and rely much more on social processes for information, inspiration and reassurance. Accordingly, opportunities for adoption are likely to be created more and more by word-of-mouth and visibility within social circles. In that regard, strengthening brand reputations would help establish trust in these technologies, whilst encouraging peer-to-peer recommendations would help provide a platform for the sharing of personal experiences to assist potential RET users in such complex purchase decisions.

Finally, RET adoption entails significant behavioural change. Charging an EV, for instance, requires a vastly different approach to vehicle management than filling a conventional car at a gas station.

Moreover, the early majority would avoid extensive research to understand the nuances of EV charging and required changes in their driving behaviours. Therefore, rather than more charging facilities, the average EV owner may now seek a charging system that is simpler to navigate such that the practicalities of ownership are made significantly easier. Thus, once capability and opportunity are catered for, a consumer's *intent* to purchase may depend on the range of accompanying supports available to ease their transition to new technologies. Again, peer effects will play a prominent role for the early majority as most purchase decisions at this stage are based on the trust established from the decision-making of previous adopter segments. Moreover, social status may now be determined by being one of many RET owners rather than being one of few, which seemed to be a driving force in the niche innovator and early adopter groups. Hence, the spread of influence may well overtake the spread of knowledge when customers seek a low risk buying experience. Overall, along with policies that ease economic constraints and resolve accessibility issues for a diverse customer base, the transition to RETs must be made as easy as possible to strengthen the intent to purchase and prevent push back.

5. Conclusion

This work has used a novel dataset derived from primary survey data, and secondary census and geo-location data to examine RET adoption at a granular spatial scale in Ireland. We defined RET users as residential consumers who own either an EV, solar PVs, or a heat pump in their homes. We developed RET adoption models based on machine learning approaches to predict and map both current and potential adoption patterns. Current adoption was based on the characteristics of current Irish RET adopters whilst potential adoption included, in addition, the attributes of survey respondents who do not currently own an RET but indicated a strong interest in purchasing RETs in the near future. Our methodological framework was particularly valid and useful as we had an imbalanced dataset to start with, which would normally produce biased results without suitable modifications such as oversampling and the application of appropriate predictive tools that can handle such computational adjustments. However, since every machine learning technique has its unique strengths and weaknesses, we tested a combination of approaches to analyse our data. The *k*-nearest neighbour and the random forest models were deemed the most informative for our context. Importantly, both *k*-nearest neighbours and random forests are easily interpretable, making these algorithms most suitable in policy settings. Based on the superior performance metrics of the random forest approach, we chose this model for our final predictive task of spatial analysis. The overarching aim of this work is to inform the design of policy incentives to encourage uptake in low uptake regions, and to help plan for any necessary infrastructure upgrades where a mismatch of capacity and technology use is identified.

Both current and potential RET adoption were found to be influenced by several financial and non-financial factors, including utility bills, occupation, residence periods, and the presence of other RET users in one's social circle. Several behavioural features such as perceptions of hassle, technical interest, and sustainability were also identified as being key to adoption. Interestingly, social approval appeared to play an important role in potential uptake but not in current uptake, implying that the Irish RET customer base is probably in transition, moving from the early adopter to the early majority phase. Thus, although the prospects for take-up look promising, the chasm between these two stages must be bridged before EVs, solar PVs and heat pumps can reach more widespread acceptance. As RETs become cheaper and markets for technologies become more open and increasingly decentralised, the difficulties in optimising resource use, capturing market niche, and maintaining momentum can be facilitated by holistically addressing consumers' intent, capability, and opportunity to purchase. Our study lays the foundation for understanding RET adoption in Ireland, by pinpointing regions where policy incentives and infrastructure upgrades are specifically required and demonstrating that efficient policy design can help boost adoption in low-uptake regions by focussing on the key predictors identified for incentive development. Our spatial findings should also assist planners in their audit of power grids and public EV charging facilities, such that RET users can have reliable access to key infrastructure without having resources lying idle for significant periods of time. Importantly, since this work utilised consumer characteristics, other early adopter markets that lack consumer behaviour data could use this information to inform their own RET uptake management policies.

Limitations

There are several limitations to our work. Firstly, our dataset for generating predictions was derived from three separate data sources that were merged to fit the project objectives. This was a non-trivial exercise and some information may have been lost in the process such as when recategorizing variables to equalise feature levels across datasets.

Secondly, there were several granularity issues with the survey dataset that we had to contend with. The key issues were low sample sizes and the lack of geo-location information. We corrected the lack of geo-location data by merging information from external datasets, viz. the 2016 census database and a bank of addresses from the Irish GeoDirectory. Since we did not have any census information at the individual level, we used small area boundaries as our lowest level of disaggregated data. The GeoDirectory provided geographic coordinates for all 18,641 small area centroids. The characteristics of each area were defined by a measure of central tendency, the mode, that is, the highest category reported within each small area. However, no corrections could be made for low sample sizes for individual technologies and thus, we had to resort to studying overall RET uptake rather than explore the adoption patterns for specific technologies.

Thirdly, any predictive modelling has several limitations. When projecting into the future with past data, the conditions prevailing such as those surrounding awareness and attitudes are implicitly assumed to continue, which is highly unlikely beyond a certain timeframe. Hence, to guarantee a reasonable level of accuracy, our predictions relate only to the very near future, not beyond 5 years at most. Moreover, it is impossible to predict human behaviour fully without knowing all possible reasons to adopt RETs. Also, the occurrence of external shocks such as extreme and rare events are inherently unpredictable. An example of this is the recent impact of the SARS-CoV-2 pandemic on carbon emissions and behaviour. These events could potentially disrupt the predicted adoption patterns. Our models are, thus, not meant to be a general prediction method and our predictions are primarily useful as a baseline for policy development for this specific case study if the limitations of modelling are acknowledged and understood. The models would be best used in conjunction with up-to-date real-world adoption data; it is crucial to review and revise the model-based classifications as and when new data become available.

Fourthly, our analysis reveals the presence of several RET adopter clusters defined by geographic location. However, clusters could also be experienced in time influenced by common exposures. Thus, for a more comprehensive picture of the regional trends in RET take-up, a spatio-temporal modelling approach is needed.

Finally, RET take-up will be fundamental to, and yet only a fragment of, any large-scale climate response. Any micro-level strategies discussed here may have insufficient impact alone and must be considered only as part of a larger environmental policy package. Overall, nudging individual behavioural changes through market mechanisms such as financial signals and social cues must be coupled with wider State-led macro-measures that impact large-scale operations such as power generation and technology development to have any measurable impact on global carbon emissions.

Further work

Our predictive modelling should help utility companies, manufacturers, planners, and policy makers estimate and visualise the spread of RET uptake in Ireland in the very near future. Consequently, this should help them formulate suitable policies to boost utilisation in low uptake areas and devise strategies to revamp infrastructure in high uptake areas.

Further work will include the collection of more detailed survey data with larger sample sizes for EV, solar PV and heat pump adopters such that similar analyses could be performed for individual technologies. The updated survey would require collection of locational data such as respondents' postal codes such that a more accurate representation of the spatial patterns of adoption and the relevance of socio-demographics can be attained. These would eliminate the need to use average regional values to generate predictions, thereby providing a more accurate picture of the heterogeneity in uptake such as the Irish rural-urban divide alluded to earlier. Furthermore, with updated survey data, a more detailed policy analysis examining adoption scenarios with varying levels of financial and non-financial measures could be carried out.

However, until any new data becomes available, our plan for next steps is two-fold. We will first map our current predictions against real-world maps depicting the location of existing power grids, EV charging stations, and the housing stock, whereby infrastructural issues such as the possibility of power failure due to high regional uptake, the impact of excessive queuing times at public charge points on regional take-up, and potential low adoption due to market failures in rental accommodation, can be explored. We will also attempt to combine our spatial analysis with a temporal diffusion model such that spatio-temporal predictions can be calculated for RET uptake in Ireland.

Acknowledgements

This work forms part of a collaborative project on the impact of renewable energy technology adoption on the energy transition and electricity network planning in Ireland. The author is grateful to the electricity distribution system operator, ESB Networks, for funding the national survey, and market research agency, Amárach Research, for assisting in the survey data collection. Informed consent was obtained from all participants involved in this research and the full study received approval from the UCD Office of Research Ethics before data collection commenced. The author also thanks researchers at the UCD Energy Institute for providing helpful feedback on previous drafts of this research. This publication is based on research supported in part by the Irish Research Council with additional funding provided by the UCD Energy Institute. There are no known conflicts of interest.

Data Availability

Unfortunately, we do not have permission to make the primary data associated with this research available due to data confidentiality reasons.

References

[1] EU. Directive (EU) 2018/2001 of the European Parliament and of the Council on the promotion of the use of energy from renewable sources. Off J Eur Union 2018;2018:82–209.

[2] European Commission. Renewable energy in europe. 2020. https://doi.org/10.1080/01425919008941479.

[3] EirGrid. Tomorrow's Energy Scenarios. Ireland: 2019.

[4] Holland M. Norway Hits 64.4% EV Market Share In January, Up 24% Year On Year. CleanTechnica 2020. https://cleantechnica.com/2020/02/05/norway-hits-64-4-ev-market-share-in-january-up-24-year-on-year/ (accessed August 29, 2020).

[5] Road Traffic Information Council. Car sales Registration Statistics. OFV Stat n.d. https://statistikk.ofv.no/ (accessed August 29, 2020).

[6] Wirth H. Recent facts about photovoltaics in Germany. 2020.

[7] IEA. Heat Pumps 2020. https://www.iea.org/reports/heat-pumps (accessed August 29, 2020).

[8] HPA. Delivering net zero: a roadmap for the role of heat pumps. Heat Pump Assoc 2019.

[9] European Heat Pump Association. Growing for good? The European Heat Pump Market -Status and outlook 2017.

[10] Shivakumar A, Dobbins A, Fahl U, Singh A. Drivers of renewable energy deployment in the EU: An analysis of past trends and projections. Energy Strateg Rev 2019;26:100402. https://doi.org/10.1016/j.esr.2019.100402.

[11] European Commission. Progress reports: Energy 2020. https://ec.europa.eu/energy/topics/renewable-energy/progress-reports_en (accessed August 29, 2020).

[12] SEAI. Renewable Energy in Ireland. 2020.

[13] Claudy MC, Peterson M, O'Driscoll A. Understanding the Attitude-Behavior Gap for Renewable Energy Systems Using Behavioral Reasoning Theory. J Macromarketing 2013;33:273–87. https://doi.org/10.1177/0276146713481605.

[14] Mukherjee SC, Ryan L. Factors influencing early battery electric vehicle adoption in Ireland. Renew Sustain Energy Rev 2020;118:109504. https://doi.org/10.1016/j.rser.2019.109504.

[15] Fernandes BS. What are the main barriers that are preventing Ireland from achieving the EU Renewable Energy Directive targets , considering its renewable resources ? Dublin Business School, 2019.

[16] Meles T, Ryan L, Mukherjee SC. Preferences for Renewable Home Heating: A Choice Experiment Study of Heat Pump System in Ireland. 24th Annu. Conf. Eur. Assoc. Environ. Resour. Econ., Manchester: 2019. https://doi.org/http://hdl.handle.net/10197/11467.

[17] Ghoddusi H, Creamer GG, Rafizadeh N. Machine learning in energy economics and finance: A review. Energy Econ 2019;81:709–27. https://doi.org/10.1016/j.eneco.2019.05.006.

[18] Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Leanpub; 2020.

[19] Walker S, Khan W, Katic K, Maassen W, Zeiler W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. Energy Build 2020;209:109705. https://doi.org/10.1016/j.enbuild.2019.109705.

[20] Al-dosary NMN, Al-hamed SA, Aboukarima AM. K-NEAREST NEIGHBORS METHOD FOR PREDICTION OF FUEL CONSUMPTION IN TRACTOR – CHISEL PLOW SYSTEMS Most important farm operations require a significant amount of energy , and this consumes a major portion of the farm ' s budget . Consequently , analyzing the . Eng Agrícola 2019;4430:729–36.

[21] Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build 2012;49:560–7. https://doi.org/10.1016/j.enbuild.2012.03.003.

[22] Chen YT, Piedad E, Kuo CC. Energy consumption load forecasting using a level-based random forest classifier. Symmetry (Basel) 2019;11:1–9. https://doi.org/10.3390/sym11080956.

[23] Chang HC, Kuo CC, Chen YT, Wu W Bin, Piedad EJ. Energy consumption level prediction based on classification approach with machine learning technique. Proc. World Congr. New Technol., Madrid: 2018, p. 1–8. https://doi.org/10.11159/icert18.108.

[24] Vialetto G, Noro M. Enhancement of a short-term forecasting method based on clustering and KNN: Application to an industrial facility powered by a cogenerator. Energies 2019;12. https://doi.org/10.3390/en12234407.

[25] Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S. Random Forest based hourly building energy prediction. Energy Build 2018;171:11–25. https://doi.org/10.1016/j.enbuild.2018.04.008.

[26] Zhang W, Robinson C, Guhathakurta S, Garikapati VM, Dilkina B, Brown MA, et al. Estimating residential energy consumption in metropolitan areas: A microsimulation approach. Energy 2018;155:162–73. https://doi.org/10.1016/j.energy.2018.04.161.

[27] Raihanian Mashhadi A, Behdad S. Discriminant effects of consumer electronics use-phase attributes on household energy prediction. Energy Policy 2018;118:346–55. https://doi.org/10.1016/j.enpol.2018.03.059.

[28] Ahmad MW, Reynolds J, Rezgui Y. Random Forests and Artificial Neural Network for Predicting Daylight Illuminance and Energy Consumption BRE Centre for Sustainable Engineering , School of Engineering , AhmadM3@Cardiff.ac.uk ; 2 HippolyteJ@Cardiff.ac.uk ; Related work. Build. Simul., San Francisco: 2017, p. 1–7.

[29] Zagar A, Grolinger K, Capretz M, Seewald L. Energy cost forecasting for event venues. IEEE Electr Power Energy Conf Smarter Resilient Power Syst EPEC 2015 2015:220–6. https://doi.org/10.1109/EPEC.2015.7379953.

[30] Durica M, Frnda J, Svabova L. Decision tree based model of business failure prediction for Polish companies. Oeconomia Copernicana 2019;10:453–69. https://doi.org/10.24136/oc.2019.022.

[31] Dogah KE, Premaratne G. Sectoral exposure of financial markets to oil risk factors in BRICS countries. Energy Econ 2018;76:228–56. https://doi.org/10.1016/j.eneco.2018.09.014.

[32] Sun J, Lang J, Fujita H, Li H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. Inf Sci (Ny) 2018;425:76–91. https://doi.org/10.1016/j.ins.2017.10.017.

[33] Liu C, Hu Z, Li Y, Liu S. Forecasting copper prices by decision tree learning. Resour Policy 2017;52:427–34. https://doi.org/10.1016/j.resourpol.2017.05.007.

[34] Dey PK. Project risk management using multiple criteria decision-making technique and decision tree analysis: A case study of Indian oil refinery. Prod Plan Control 2012;23:903–21. https://doi.org/10.1080/09537287.2011.586379.

[35] Fan GZ, Ong SE, Koh HC. Determinants of house price: A decision tree approach. Urban Stud 2006;43:2301–16. https://doi.org/10.1080/00420980600990928.

[36] Fernández Rodríguez F, Sosvilla Rivero SJ, Andrada Félix J. Nearest-Neighbour Predictions in Foreign Exchange Markets. In: Chen SH, Wang PP, editors. Comput. Intell. Econ. Financ. Adv. Inf. Process., Berlin, Heidelberg: Springer; 2004. https://doi.org/10.2139/ssrn.300404.

[37] Aparicio T, Pozo E, Saura D. The nearest neighbour method as a test for detecting complex dynamics in financial series. An empirical application. Appl Financ Econ 2002;12:517–25. https://doi.org/10.1080/09603100010007986.

[38] Álvarez-Díaz M, Mateu-sbert J. Forecasting daily air arrivals in Mallorca Island using nearest neighbour methods. Tour Econ 2011;17:191–208. https://doi.org/10.5367/te.2011.0022.

[39] Biau O, D'Elia A. Euro Area GDP Forecasting Using Large Survey Datasets: A Random Forest Approach. Euroindicators Work Pap 2011/002 2011. https://doi.org/10.2901/1977-3331.2011.002.

[40] Alvarez-Diaz M. Exchange rates forecasting: Local or global methods? Appl Econ 2008;40:1969–84. https://doi.org/10.1080/00036840600905308.

[41] Kumar M, M. T. Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. SSRN Electron J 2006:1–16. https://doi.org/10.2139/ssrn.876544.

[42] Pedersen CS, Satchell SE. Evaluating the Performance of Nearest Neighbour Algorithms when Forecasting US Industry Returns. Financ Econom Res Cent 2000:1–20.

[43] Mercadier M, Lardy JP. Credit spread approximation and improvement using random forest regression. Eur J Oper Res 2019;277:351–65. https://doi.org/10.1016/j.ejor.2019.02.005.

[44] Antipov EA, Pokryshevskaya EB. Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. Expert Syst Appl 2012;39:1772–8. https://doi.org/10.1016/j.eswa.2011.08.077.

[45] Bozsik J, Körmendi G. Decision tree-based credit decision support system. 3rd IEEE Int Symp Logist Ind Informatics, Proc 2011:189–93. https://doi.org/10.1109/LINDI.2011.6031145.

[46] Liu C, Chan Y, Alam Kazmi SH, Fu H. Financial Fraud Detection Model: Based on Random Forest. Int J Econ Financ 2015;7:178–88. https://doi.org/10.5539/ijef.v7n7p178.

[47] Sohn SY, Kim JW. Decision tree-based technology credit scoring for start-up firms: Korean case. Expert Syst Appl 2012;39:4007–12. https://doi.org/10.1016/j.eswa.2011.09.075.

[48] De Reyck B, Degraeve Z, Vandenborre R. Project options valuation with net present value and decision tree analysis. Eur J Oper Res 2008;184:341–55. https://doi.org/10.1016/j.ejor.2006.07.047.

[49] Lin W, Wu Z, Lin L, Wen A, Li JIN. An Ensemble Random Forest Algorithm for Insurance Big Data Analysis. IEEE Access 2017;5.

[50] Kusuma PYC, Sumpeno S, Wibawa AD. Social media analysis of BPS data availability in economics using decision tree method. Proc - 2016 1st Int Conf Inf Technol Inf Syst Electr Eng ICITISEE 2016 2016:148–53. https://doi.org/10.1109/ICITISEE.2016.7803064.

[51] Delen D, Kuzey C, Uyar A. Measuring firm performance using financial ratios: A decision tree approach. Expert Syst Appl 2013;40:3970–83. https://doi.org/10.1016/j.eswa.2013.01.012.

[52] Wu MC, Lin SY, Lin CH. An effective application of decision tree to stock trading. Expert Syst Appl 2006;31:270–4. https://doi.org/10.1016/j.eswa.2005.09.026.

[53] Long W, Song L, Cui L. Relationship between Capital Operation and Market Value Management of Listed Companies Based on Random Forest Algorithm. Procedia Comput Sci 2017;108:1271–80. https://doi.org/10.1016/j.procs.2017.05.111.

[54] Jiang X, Abdel-Aty M, Hu J, Lee J. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing 2016;181:53–63. https://doi.org/10.1016/j.neucom.2015.08.097.

[55] Rogers EM. Diffusion of innovations. Free Press of Glencoe; 1962. https://doi.org/10.4324/9780203710753-35.

[56] Chattopadhyay Mukherjee S, Hadush Meles T, Ryan L, Healy S, Mooney R, Sharpe L, et al. Renewable Energy Technology Uptake: Public Preferences and Policy Design in Early Adoption Sanghamitra. 2020.

[57] CSO. Census 2016 Small Area Population Statistics - CSO - Central Statistics Office. CSO 2017. https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/ (accessed August 29, 2020).

[58] An Post, Ordnance Survey Ireland. GeoAddress. GeoDirectory n.d. https://www.geodirectory.ie/products-services/geoaddress (accessed August 29, 2020).

[59] R Core Team. R: A language and environment for statistical computing 2020.

[60] Cover TM, Hart PE. Nearest Neighbor Pattern Classification. IEEE Trans Inf Theory 1967;I:771–771. https://doi.org/10.1007/978-0-387-35973-1_862.

[61] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan US. When is nearest neighbor meaningful.pdf. Int Conf Database Theory 1999:217–35.

[62] Pestov V. Is the k-NN classifier in high dimensions affected by the curse of dimensionality? Comput Math with Appl 2013;65:1427–37. https://doi.org/10.1016/j.camwa.2012.09.011.

[63] Kotu V, Deshpande B. Predictive Analytics and Data Mining. First. Morgan Kaufmann; 2014. https://doi.org/10.1016/B978-0-12-801460-8.00004-5.

[64] Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106. https://doi.org/10.1007/bf00116251.

[65] Breiman L. Random forests. Mach Learn 2001;45:5–32. https://doi.org/10.1201/9780367816377-11.

[66] ESRI. ArcGIS Online 2020.

[67] NGA. World Geodetic System (WGS84). Natl Geospatial-Intelligence Agency 1984. https://www.nga.mil/search/index.html?query=wgs (accessed September 12, 2020).

[68] Lyndhurst B. The diffusion of environmental behaviours; the role of influential individuals in social networks. London: 2009.

[69] Allan C, Jaffe AB, Sin I, Allan C. Diffusion of Green Technology: A Survey. 2014.

[70] Zeng Y, Dong P, Shi Y, Wang L, Li Y. Analyzing the co-evolution of green technology diffusion and consumers ' pro-environmental attitudes : An agent-based model. J Clean Prod 2020;256:120384. https://doi.org/10.1016/j.jclepro.2020.120384.

[71] OECD. Behavioural insights. Organ Econ Co-Operation Dev 2020. https://www.oecd.org/gov/regulatory-policy/behavioural-insights.htm (accessed September 12, 2020).

[72] Purcell K. Applying Behavioural Economics in Irish Policy. 2016.

[73] Comin D, Dmitriev M, Rossi-Hansberg E. The Spatial Diffusion of Technology. Natl Bur Econ Res 2012. https://doi.org/10.3386/w18534.

[74] Henderson V. The Global Distribution of Economic Activity: Nature, History, and the Role of Trade. Q J Econ 2018.

[75] CSO. Life in 1916 Ireland: Stories from statistics. Cent Stat Off 2020. https://www.cso.ie/en/releasesandpublications/ep/p-1916/1916irl/people/population/ (accessed August 29, 2020).

[76] CSO. Regional SDGs Ireland 2017. Cent Stat Off 2020. https://www.cso.ie/en/releasesandpublications/ep/p-rsdgi/regionalsdgsireland2017/ (accessed August 29, 2020).

[77] World Population Review. Population of Cities in Ireland (2020). World Popul Rev 2020.

[78] Chen C, Liaw A, Breiman L. Using Random Forest to Learn Imbalanced Data. 2004.

[79] Analytics Vidhya. Practical Guide to deal with Imbalanced Classification Problems in R. Anal Vidhya 2016. https://www.analyticsvidhya.com/blog/2016/03/practical-guide-deal-imbalanced-classification-problems/ (accessed August 29, 2020).

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP20/03 Vincent Hogan and Patrick Massey: 'Professional Rugby on the Celtic Fringe' January 2020
WP20/04 Sanghamitra Chattopadhyay Mukherjee, Séin Healy, Tensay Hadush Meles, Lisa Ryan, Robert Mooney, Lindsay Sharpe and Paul Hayes: 'Renewable Energy Technology Uptake: Public Preferences and Policy Design in Early Adoption' February 2020
WP20/05 Ronald B Davies and James R Markusen: 'The Structure of Multinational Firms' International Activities' February 2020
WP20/06 Morgan Kelly: 'Direct Standard Errors for Regressions with Spatially Autocorrelated Residuals' March 2020
WP20/07 Anna Bayona and Oana Peia: 'Financial Contagion and the Wealth Effect: An Experimental Study' March 2020
WP20/08 Orla Doyle: 'Can Early Intervention have a Sustained Effect on Human Capital?' April 2020
WP20/09 Supriya Kapoor and Oana Peia: 'The Impact of Quantitative Easing on Liquidity Creation' April 2020
WP20/10 Judith M Delaney and Paul J Devereux: 'Choosing Differently? College Application Behaviour and the Persistence of Educational Advantage' May 2020
WP20/11 Jonathan Briody: 'Parental Unemployment During the Great Recession and Childhood Adiposity' May 2020
WP20/12 Tensay Meles, Lisa Ryan, and Joe Wheatley: 'COVID-19 and EU Climate Targets: Going Further with Less?' May 2020
WP20/13 Xidong Guo and Sarah Parlane: 'Addressing Private Practice in Public Hospitals' May 2020
WP20/14 Patrick Massey and Moore McDowell: 'EU Competition Law: An Unaffordable Luxury in Times of Crisis?' May 2020
WP20/15 Judith M Delaney and Paul J Devereux: 'How Gender and Prior Disadvantage predict performance in College' May 2020
WP20/16 Morgan Kelly, Joel Mokyr and Cormac Ó Gráda: 'The Mechanics of the Industrial Revolution' June 2020
WP20/17 Morgan Kelly and Cormac Ó Gráda: 'Connecting the Scientific and Industrial Revolutions: The Role of Practical Mathematics' June 2020
WP20/18 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' June 2020
WP20/19 Ronald B Davies, Yutao Han, Kate Hynes, and Yong Wang: 'Competition in Taxes and IPR' June 2020
WP20/20 David Madden: 'The Socioeconomic Gradient of Cognitive Test Scores: Evidence from Two Cohorts of Irish Children' June 2020
WP20/21 Deirdre Coy and Orla Doyle: 'Should Early Health Investments Work? Evidence from an RCT of a Home Visiting Programme' July 2020
WP20/22 Cormac Ó Gráda: 'On Plague and Ebola in a Time of COVID-19' August 2020
WP20/23 Morgan Kelly: 'Understanding Persistence' September 2020
WP20/24 Jeff Concannon and Benjamin Elsner: 'Immigration and Redistribution' September 2020
WP20/25 Diego Zambiasi: 'Drugs on the Web, Crime in the Streets - The Impact of Dark Web Marketplaces on Street Crime' September 2020
WP20/26 Paul Hayes, Séin Healy, Tensay Meles, Robert Mooney, Sanghamitra Mukherjee, Lisa Ryan and Lindsay Sharpe: 'Attitudes to Renewable Energy Technologies: Driving Change in Early Adopter Markets' September 2020

UCD Centre for Economic Research
Email economics@ucd.ie