Research topic:

Third Energy Revolution

Regulatory Intervention, Market Forces and Disruptive Innovation Factors

Submitted by:

Gil Barnea

Department of Public Policy

Chapter XX – Support schemes \ multimethod approach for renewable energies in worldwide comparison

Contents

[Table of Figures 3](#_Toc15750649)

[Index of Tables 3](#_Toc15750650)

[Abbreviations 4](#_Toc15750651)

[1. Introduction 5](#_Toc15750652)

[2. Literature Review 8](#_Toc15750653)

[3. Methodology 9](#_Toc15750654)

[3.1. Correlation Matrix 9](#_Toc15750655)

[3.2. Multiple Regression 10](#_Toc15750656)

[3.3. Support Vector Machine 10](#_Toc15750657)

[3.4. CHAID Classification Tree 13](#_Toc15750658)

[3.5. Bayesian belief network 14](#_Toc15750659)

[4. Data 15](#_Toc15750660)

[4.1. Policy instruments (REN21) 15](#_Toc15750661)

[4.2. Fiscal and market indicators 16](#_Toc15750662)

[4.3. Governance Indicators (World Bank) 16](#_Toc15750663)

[4.4. Development indicators (Diverse Sources) 17](#_Toc15750664)

[4.5. RISE Regulatory Indicators (World Bank) 18](#_Toc15750665)

[5. Results 19](#_Toc15750666)

[5.1. Correlation Matrix 19](#_Toc15750667)

[5.2. Multiple Regression 21](#_Toc15750668)

[5.3. SVM analysis 21](#_Toc15750669)

[5.3.1. results 23](#_Toc15750670)

[5.4. CHAID Classificaion Trees 24](#_Toc15750671)

[5.4.1. Design of the classification\ Decision Trees 24](#_Toc15750672)

[5.4.2. Results 25](#_Toc15750673)

[5.4.3. Disscussion 30](#_Toc15750674)

[5.5. BAYESIAN Model 30](#_Toc15750675)

[5.5.1. Bayes’ Theorem 31](#_Toc15750676)

[5.5.2. Bayesian Belief Networks 31](#_Toc15750677)

[5.5.3. Methodology 32](#_Toc15750678)

[5.5.4. Data 33](#_Toc15750679)

[5.5.5. Results 35](#_Toc15750680)

[6. Discussion 37](#_Toc15750681)

[7. Conclusion 39](#_Toc15750682)

[8. Summery 40](#_Toc15750683)

[9. Bibliography 42](#_Toc15750684)

# Table of Figures

[Figure 1 multimethod research design 9](#_Toc15750685)

[Figure 2 Support vectors 11](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750686)

[Figure 3 Hyperplanes 11](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750687)

[Figure 4 Kernel function 12](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750688)

[Figure 6 131 countries Classification Tree 28](#_Toc15750689)

[Figure 7 64 countries Classification Tree 30](#_Toc15750690)

[Figure 8 Bayes' Theorem sample (Wikipedia 2011) 31](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750691)

[Figure 9 BBN model (The case of Greece as an example) 34](#_Toc15750692)

[Figure 10 VRE penetration Conditional Probability Table (CRT) example 35](#_Toc15750693)

[Figure 11 Simulation E 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750694)

[Figure 12 Simulation F 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750695)

[Figure 13 Simulation D 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750696)

[Figure 14 Calibration 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750697)

[Figure 15 Simulation E 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750698)

[Figure 16 Simulation H 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750699)

[Figure 17 Simulation G 36](file:///C:\Users\Gil's%20Surface%20PC\Dropbox\PHD\PhD%20Dissertation\Support%20schemes%20chapter\Dissertation%20Support%20schemes%20Chapter%20XXX.docx#_Toc15750700)

# Index of Tables

[Table 2 Relationship between World Bank governmental indicators 20](#_Toc15750701)

[Table 3 Variables correlation to VRE 20](#_Toc15750702)

[Table 1 RISE indicators correlation to VRE 21](#_Toc15750703)

[Table 4 SVM Results 24](#_Toc15750704)

[Table 5 CHAID well-classified observations 25](#_Toc15750705)

[Table 6 131 countries Classification tree structure 26](#_Toc15750706)

[Table 7 64 countries Classification tree structure 29](#_Toc15750707)

[Table 8 Raw data starting point 34](#_Toc15750708)

[Table 9 BBNs Indicators weight 34](#_Toc15750709)

[Table 10 Results table 36](#_Toc15750710)

[Table 11 Summery of outputs 37](#_Toc15750711)

# Abbreviations

VRE Variable Renewable Energy

FIT Feed-in tariff

NET Net metering/net billing

T-REC Tradable Renewable Energy Certificate

CS Capital subsidy, grant, or rebate

INV Investment or production tax credits

PI Public Investment, loans, or grants

RQ Regulatory Quality

EV Electric Vehicle

CRT Conditional Probability Table for Netica model

SVM Support Vector Machine

BBN Bayesian Belief Networks

CHAID Chi-squared Automatic Interaction Detector

WGI World-Bank Governance Indicators

RISE Regulatory Indicators for Sustainable Energy

### Introduction

Renewable energy offers ecological benefits and energy independence to developed as well as developing countries (World-Bank, 2018a). Surprisingly, the expansion of wind and solar energy capacity varies tremendously between different countries. While some countries have changed their energy mix to a remarkable extent, others have not yet unlocked their potential of renewable energies (Hales, 2018).

Many countries have declared their intention to encourage VRE (Variable Renewable Energy), and are actively promote legislation, support, and budgeting of the field, but did not reach the desired outcomes. In addition, the vast difference in the rate of adoption of renewable energies between countries requires a thorough examination of why some countries motivate renewable energy, and other countries do not. The main research question of this thesis is:

***Do governmental regulations and supporting policy plays a vital role in promoting new markets that require extensive investment and infrastructure in the inception and initial rollout phase?***

In order to answer the hypotheses that governmental regulations and supporting policy are essential to promote technology that requires extensive investment and infrastructure, the chapter examines the specific question about the critical and influential regulations that create the market of renewable energies.The goal of this chapter is to recognize the factors influencing and differences between leading, significant, and influential countries in this revolution and lagging ones.

According world bank, setting national VRE targets signals the market that the government is committed to developing renewable energy. However, a target itself is not enough, as even the most ambitious goals are meaningless without a clear understanding of what needs to be done or a realistic strategy for implementation (Banerjee, Sinton, Moreno, Primiani, & Seong, 2016, p. 56). Therefore, the government requires to promote a strategy or long-term action plan; to state the required investments, infrastructures and regulations needed to meet the target. These actions are typically designed to minimize costs while ensuring reliable and sufficient lectricity supply.

One can analyze the impact of each regulatory or market instrument individually at the national level, assuming no cross-effect relationship between tools or structural issues and external disputes that may affect the penetration progression. Linking all the factors from a global view, and hence a more efficient and proper construction of this resource-intensive market is substantial at the country level.

The Paris Agreement signed in 2015 to combat global warming and to accelerate and intensify the investments and actions needed for a low carbon prospect. Today, several years after, it is an excellent inspection point for conducting systematic analysis and quantitative comparison among countries in order to understand the factors that directly and effectively influence the penetration of VRE.

This chapter analyzes 131 countries and 31 indicators in various methods in order to understand the influence network, correlations, and to predict the penetration of VRE. The 31 variables examined in this study creates “overidentified model” and cover a wide range of areas - regulation, market, and natural resources. Even thou, the possibility that other factors are influencing the progression can never be ruled out.

The research uses a multi-methods approach (distinction from mixed methods that combines qualitative and quantitative methods (Clark, 2007)). Using Pearson correlation matrix, Multivariate regression analysis, Support Vector Machine (SVM), a machine learning method, Chi-squared Automatic Interaction Detector (CHAID) classifier, and Bayesian networks probabilistic graphical model, enable us to merge between the advantages of each method. These methods have been widely adopted in the field of medicine and exact science but not in policy studies. As far as I know, based on in-depth research in leading academic search engines as of Google Scholar and Microsoft Academic, no previous research has investigated these issues with the following methods. All the statistical assessments were performed using XLSTAT software.

Crucially, the dividing intersects between countries holding and lacking renewables does not necessarily run alongside their economic development. This chapter thus, asks for factors to explain varying patterns of renewable energy potential exploitation. In addition to structural conditions (economic development, government capacity, regime type). The study focus on financial and environmental indicators, funding instruments for renewable energy development, market-oriented tools (quota systems, tendering), and regulatory and direct public interventions (feed-in tariffs, public investment).

The results of this research will enable a more efficient and fast progression at the country level in the future. With the tools utilized in this chapter, one can now accomplish a high-level assessment on VRE's penetration rate for the rest of the world, about 60 countries, for which we did not have data. In addition, these different tools can be used as a model for countries with a low penetration rate, to recalculate the steps required in the future to introduce VRE effectively or to design an efficient process for increasing and accelerating VRE implementation.

### ****Literature Review****

Part of the Ph.D. proposal. Will be added later.

### ****Methodology****

Multimethod research approach includes the use of more than one method of data collection or research method in a research study. In this work, I analyzed and investigated the same data with various exploration tools in order to get as many conclusions as possible from the same dataset. Using Pearson correlation matrix, Multivariate regression analysis, Support Vector Machine (SVM), a machine learning method, Chi-squared Automatic Interaction Detector (CHAID) classifier, and Bayesian networks probabilistic graphical model, enable us to merge between the advantages of each method. Figure 1 describes the flow of the following research graphically. The data in this work consists of two datasets – 131 counties from REN21 and 105 countries from RISE\ World-Bank. For the RISE dataset, this analysis used multiple regression, and for the REN21 dataset, five different diagnostic tools were used.

Figure 1 multimethod research design

### Correlation Matrix

A correlation matrix is a table presenting correlation coefficients between different variables. Each cell in the table shows the correlation between two of the tested variables. A correlation matrix is a well-sited way to summarize the relationship between variables, as an input into a more advanced analysis (Bock, 2019). In our case, Pearson's correlation coefficient (r) **measures the strength of the connection** between the two variables (only for linear connection). **Correlation matrix is calculated to review a large number of variables where the goal is to see patterns, to arrange data entered into other analyses, and as a diagnostic tool.**

In order to have data classification and overview of the relationship between the policy instruments, a correlation matrix was performed with all data obtained.

### Multiple Regression

Linear regression is a quantitative method to investigate correlations among dependent and independent variables when a study requires an analysis of more than two variables. This study investigates the connection between variables related to GDP, population, education, demography, and radiation (as **independent variables**) and implementation of solar energy in states of India (as the **dependent variable**).

The primary regression models include unknown parameters (β), independent variables (X), and the dependent variable (Y). The regression model specifies the variation of the dependent variable (Y) or variables as a function of independent variables (X) and unknown parameters (β) - Y ≈ *f (X, β)* (Addinsoft, 2019)*.*

The regression equation utilizes a best fit straight (regression) line to predict the ‘y’ values if the value of ‘x’ is given, and both ‘y’ and ‘x’ are the two sets of measures of a sample size of ‘n.’ The formula for the regression equation would be in the form: *y = a + bx*. The statistical significance level that was considered was a p-value less than or equal to 0.05.

### Support Vector Machine

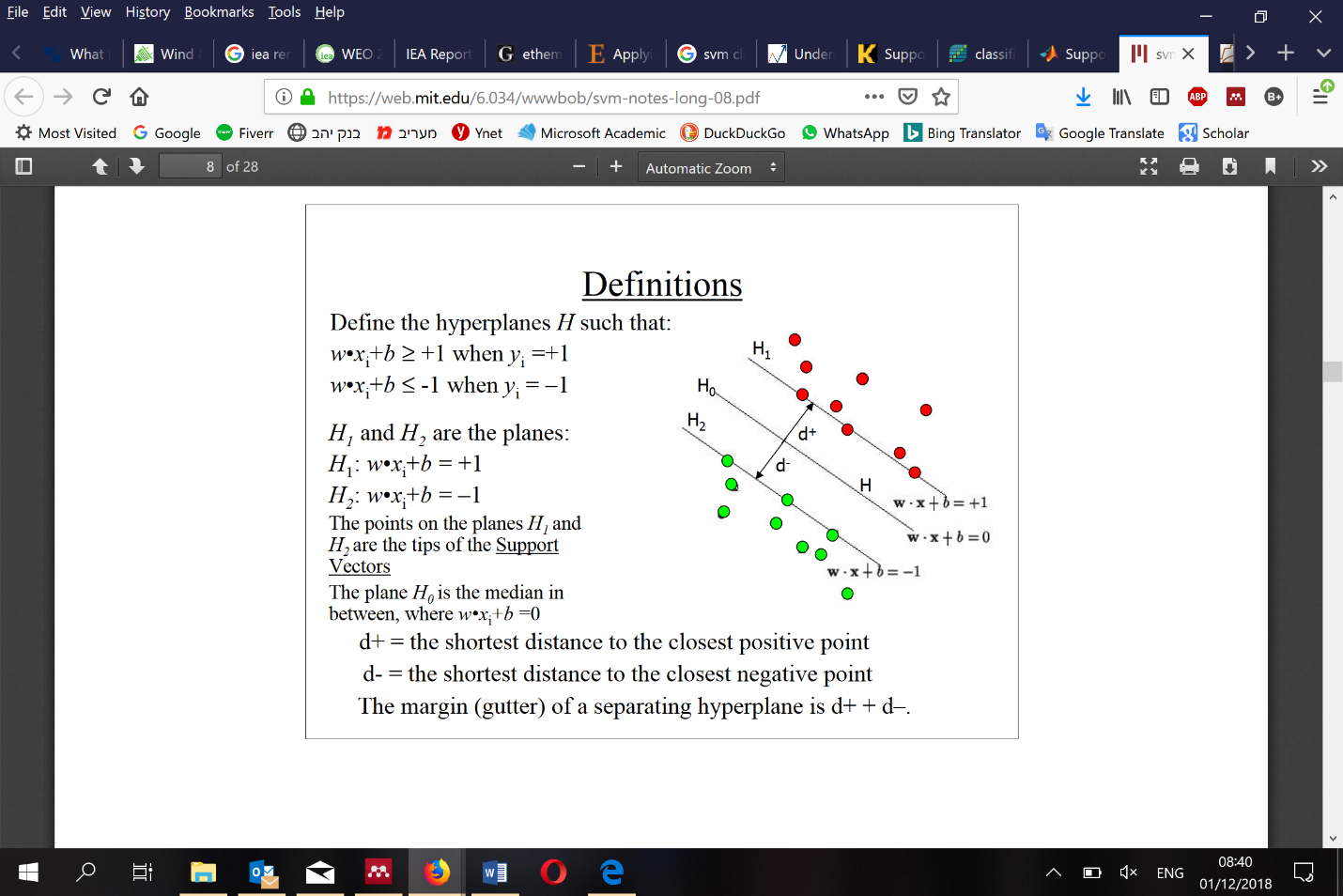
Support Vector Machines (SVM) aim at finding the separation between two classes of objects (the larger the separation, the more reliable the classification). SVM is a machine learning classification that uses a hypothesis space of linear functions in a high dimensional space, trained with optimization algorithms that implement a learning bias derived from the theory of statistical learning. The algorithm selects a hyperplane that separates the observations into two distinct classes to maximize the closest distance between the hyperplane and the observation of the training set.

Figure 2 Support vectors

The introduction of the new method was in 1964 in the context of the statistical learning theory (Vapnik & Chervonenkis, 1964). The first implementation was with the introduction of the kernel trick (Boser, Guyon, & Vapnik, 1992) and the generalization to the non-separable case (Cortes & Vapnik, 1995). Since then, the SVM has known several expansions and gained popularity in numerous areas such as Machine Learning, optimization, neural networks, etc. SVM is a popular learning algorithm (Cherkassky & Ma, 2004).

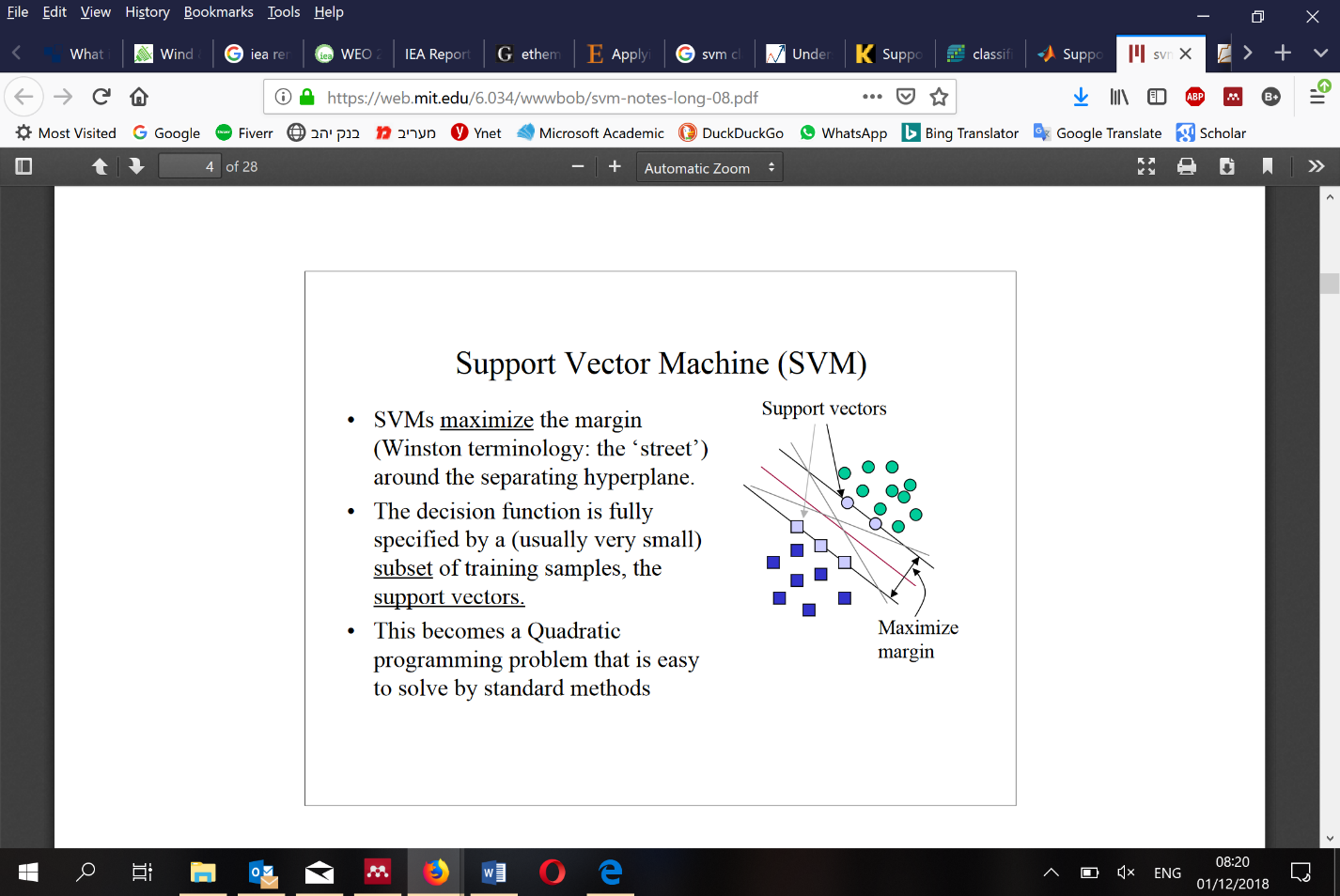
Support vectors are the points that lie adjacent to the decision surface (or hyperplanes) and are difficult to classify (Figure 2). They have a direct bearing on optimal locations in the decision space (Berwick, 1990). SVM focuses on the points that are the most challenging to differentiate, whereas other classifiers pay attention to all of the points. The support vector machine should **explicitly** identify the best separating line (Vieira, 2017). SVMs maximize the margin (as in Figure 3) around the separating hyperplane to the support vectors. The decision function is fully stated by a subset of training samples, the support vectors.

Figure 3 Hyperplanes

The SVM Defines the hyperplanes such that and are the planes:

The points on the planes and are the tips of the Support Vectors. The plane is the median in between, where

d+ = the shortest distance to the closest positive point.

d– = the shortest distance to the closest negative point.

The margin of the separating hyperplane is .

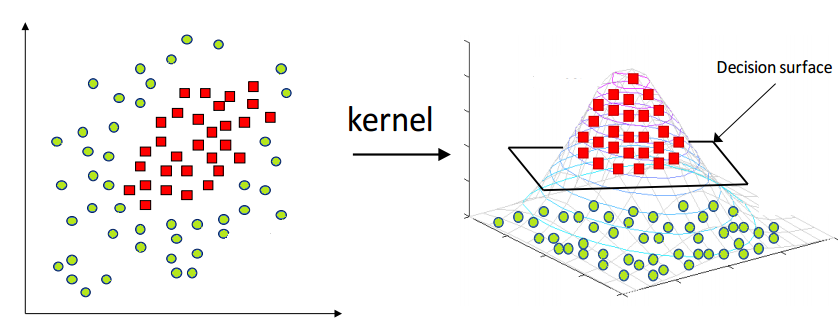
[](https://blog-c7ff.kxcdn.com/blog/wp-content/uploads/2017/02/kernel.png)The optimization algorithm generates weights so that only the support vectors determine the weights and thus the boundary. Now, finding a classifier (linear separator) with as big a margin as possible is needed. To maximize the margin, we need to minimize . With the condition that there are no data points between and . This is a constrained optimization problem that can be solved by the LaGrange multiplier method with two separate classes of data in a line (in the case of 2D data) and a plane (in the case of 3D data).

Figure 4 Kernel function

It is not always possible to use lines or planes, and one involves a nonlinear section to separate the classes; SVM handle such situations by using a Kernel function as can be seen in Figure 3 (Jain Rashmi, 2017). Transferring the data to dissimilar spaces and using a linear hyperplane allows separation of the classes.

SVM is suitable for linear and nonlinear data and works for small and high dimensional data spaces. It is effective for high-dimensional datasets because the training dataset of SVM complexity is categorized by the number of support vectors rather than dimensionality. Even if all other training examples are removed and the training repeated, we get the same optimal separating hyperplane. SVMs can work effectively on smaller training datasets and not rely on the complete data. SVMs are not suitable for larger datasets because training takes time and computational intensity. They are also less effective on noisier datasets that have overlapping modules.

XLSTAT software (version 2018.5) was used for SVM, correlation tables, CHAID decision tree analysis, and logistic regression analysis. While SVM produces probability tables, CHAID produces a tree diagram, a table giving a complete breakdown of each split, and highly detailed summaries of every step in its analysis. Using this methodology and theory enabled a significant analysis of the influential factors of RE and EVs penetration process in various countries around the world.

### CHAID Classification Tree

CHAID (**Ch**i-square **A**utomatic **I**nteraction **D**etector) analysis is an algorithm used for determining relationships between a categorical response variable and other categorical predictor variables. AID trees were first developed and published in 1963 by Morgan and Sonquist (Morgan & Sonquist, 1963). CHAID trees were later proposed by Kass (1980) and Classification And Regression Trees (C&RT) methods, was extended later by Breiman (Breiman L., Friedman, Olshen, & Stone, 1984). CHAID is useful for identifying patterns in datasets with multiple categorical variables and summarizes the data allowing easy visualization of relationships, highlighting both explanatory and predictive goals.

The strengths of this method are the use of simple graphical representations and compact formatting of natural language rules. These modeling techniques are used to explain and predict the belonging of objects to a class based on explanatory quantitative and qualitative variables. Regression trees can also build a descriptive and predictive model for a quantitative dependent variable based on explanatory quantitative and qualitative variables (XLSTAT).

CHAID was designed to process non-metric and non-ordinal data that standard multivariate analyses cannot handle; however, CHAID has limitations: data must be ordinal, nominal or interval, and not metric. No variable can have more than 15 levels, and dependent variables must be declared. CHAID will partition the sample to maximize between-group differences (variance) on this variable and cannot procedure zero values or codes that are not in sequence (Struhl Steven, 2018). The CHAID/C&RT algorithm provides more flexibility in handling data than CHAID, allowing for both continuous and categorical variables as dependent and independent variables. Using continuous dependent variables, C&RT procedures search for arrays in which the dependent variable does not vary significantly on the predictor variable. CHAID/ C&RT algorithms can handle missing values and non-continuous codes better than CHAID. With CHAID/ C&RT procedures, missing values can be left blank, and codes do not need to follow in strict sequence. These methods are used to build rules-based models to explain qualitative or quantitative dependent variables, by the most important explanatory variables, to identify groups created by the rules, and to predict the value of the dependent variable for a new observation. To complete this analysis XLSTAT software (version 2018.5) was used.

### Bayesian belief network

In recent years, Bayesian belief network (BBN) models have been applied to numerous environmental and related applications (Krause, Small, Haas, & Jaeger, 2016). A BBN is a graphical model representing the quantitative probabilistic relations between numerous events, which are related by their prior and conditional probabilities. The BBN is also a graphical structure used to describe the relationships between causes and effects in a diagram in which nodes represent events and probability description (Fenton & Neil, 2007). The network consists of the relationships among variables and nodes. Arcs between nodes indicate the causal model influence. Influence is propagated through a conditional probability table that specifies the influence between nodes.

Parameter estimation for a BBN model involves the determination of the conditional probability table for each node. This probability table may be approached using several qualitative or quantitative methods, including expert induction (Xian, 2012). In this work, the BBN could be used to analyze and predict the probability of penetration pace dependence on the results of the CHAID classification tree.

Several software applications are available for implementing BBNs (Small, 2016). I utilized Norsys Netica version 6.03 to develop decision nets, and influence diagrams model, in this study (Norsys, 2017). The software can create unidirectional causal relations and can find the best relations between the nodes to get the highest penetration percentage.

### ****Data****

For this study, I analyzed the data collected from different reliable sources. The analysis compares VRE installed capacities in 131 countries correlation to 31 different indicators taken mostly from REN21 Global Status Report (Hales, 2018), IRENA renewables capacity reports, EIA International Energy Statistics (IEA Beta, 2019), world energy council (Antonio Erias et al., 2016). I utilized secondary data from the World-Bank open data catalog, and World-Bank Governance Indicators (WGI) project (Worldbank Bank & NRGI, 2018). The indicators were divided into four different groups: Policy instruments, governance indicators, development indicators, and fiscal & market indicators.

### Policy instruments (REN21)

* Feed-in tariff/feed-in premium - A policy that typically guarantees renewable generators specified payments per unit over a fixed period. Feed-in tariff (FIT) policies also include regulations by which generators can interconnect and sell power to the grid.
* Electric utility quota obligation/RPS - (Renewable portfolio standard) Government obligation on utility companies or consumers to provide or use a minimum quota of RE share of installed capacity, or electricity generated or sold. Also called “RE standards”, “RE obligations” and “mandated market shares”.

#### Net metering/net billing - A regulated arrangement in which utility customers with on-site electricity generators can receive credits for excess generation. Customers typically receive credit equal to the retail electricity price.

#### Obligation/mandate   (Heat and transport) - A measure that requires selected partners (consumers, suppliers, generators) to reach a minimum, and often gradually increasing targets for RE, such as a stated amount of capacity, or the obligatory to use a specified RE technology. Mandates can include renewable portfolio standards (RPS) and so.

* Capital subsidy, grant, or rebate - A subsidy that includes a share of the direct capital cost of an asset. One-time payment by the government to cover a part or all of the capital cost of an investment.
* Investment or production tax credits - *Investment tax credit* -A fiscal incentive that allows investments in RE to credit part or entirely the tax obligations or income of a project developer, industry, etc. *Production tax credit -* A tax incentive that provides the owner or investor of a qualifying estate with a tax credit based on the amount of RE generated by that facility.
* Reduction in sales, energy, VAT or other taxes (Tax incentives) - Includes all tax incentives such as investment tax credits, production tax credits, and reductions in taxes on sales, energy, carbon, excise, value-added (VAT), etc.
* Energy production payment – Direct payment per unit of RE produced in a particular form.

### Fiscal and market indicators

* Public investment, loans, or grants - Governmental Financial support mechanism, often in the form of funds, grants or loans, to support the development or deployment of RE technologies.
* Tradable REC (Renewable energy certificate) - A certificate awarded to certify the generation of one unit of renewable energy. Certificates can be accumulated to meet RE obligations and also provide a trading tool among consumers and producers. Also means enabling purchases of voluntary green energy.
* Tendering - (also called auction, reverse auction or tender). A competitively purchasing mechanism. RE supply or capacity is solicited from sellers, who offer bids at the lowest price that they would be willing to accept.
* GDP growth (annual %) - Annual percentage growth rate of GDP at market prices based on constant local currency. Based on constant 2010 U.S. dollars.
* GDP per capita (current US$) - GDP per capita is gross domestic product divided by midyear population. Data are in current U.S. dollars.

### Governance Indicators (World Bank)

* Voice and accountability - The extent to which a country's citizens can participate in selecting their government, freedom of expression, association, and free media.
* Political Stability and Absence of Violence - Perceptions of the likelihood of political instability and politically motivated violence, including terrorism. Evaluation gives the country's score on the aggregate indicator, in units of a standard normal distribution.
* Government effectiveness - The quality of public services and degree of independence from political pressures of the civil service. The quality of policy construction and implementation, along with the government’s commitment to such policies.
* Regulatory quality - Regulatory quality measures government ability to formulate and implement sound policies and regulations that permit and promote private sector development.
* Rule of law - Measures the extent to which agents have confidence in and abide by the rules of society. In particular, the quality of law enforcement, police, and the courts, as well as the likelihood of crime and violence.
* Control of corruption – The extent to which public power is exercised for private gain, including insignificant and grand forms of corruption.

### Development indicators (Diverse Sources)

* Wind and Solar as share of total - VRE power systems % of total installed electricity capacity
* Population - Total population based on the definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates.
* Coal, % of world reserves - Hard coal lasting resources potential according to BP’s Statistical Review of Energy 2015.
* Coal, tones per km^2 – Hard coal reserves in tones per area
* Gas, % of world reserves - Global proved natural gas reserved
* Gas, tones per km^2 - Natural gas reserves in tones per area
* Hydro as share of total - Hydropower systems % of total installed electricity capacity
* Hydro per capita - Hydro installed capacity divided by midyear population
* Area - Countries area in KM^2
* Federal state - Countries with federal regime type. Typical for big populated countries.
* Democracy - Classification of political regimes as democracy or dictatorship(Antonio, Jennifer, & Vreeland, 2010). Classification of democracies as parliamentary, semi-presidential (mixed) and presidential. Classification of dictatorships as military, civilian and royal.

### RISE Regulatory Indicators (World Bank)

In order to asure that the results are robast, an additional set of special Regulatory Indicators for Sustainable Energy (RISE) was examined.

RISE is a set of indicators developed by a partnership between the World Bank and 18 partners, in order to help comparing policy and regulatory frameworks for sustainable energy on national level. It assesses countries’ policy and regulatory support for each of the three pillars of sustainable energy—access to modern energy, energy efficiency, and renewable energy (World-Bank, 2018a). These indicators measure regulation and policy instruments that have direct connection to VRE:

* Legal framework for renewable energy - Primary legislation, legal private ownership of generation
* Planning for renewable energy expansion - Renewable energy targets and plans, renewable energy in generation planning, and renewable energy in transmission planning.
* Incentives & regulatory support - Financial and regulatory incentives, and grid access and dispatch.
* Attributes of financial and regulatory incentives - Predictabillity and efficiency (policy-neutral), predictability and efficiency (policy-specific) and long-term sustainability.
* Network connection and use - Connection cost allocation, network usage and pricing• and renewable grid integration
* Counterparty risk - Payment risk reduction, utility creditworthiness and utility transparency and monitoring.
* Carbon pricing and monitoring - Carbon pricing mechanism, monitoring, reporting and verification (MRV) system.

### ****Results****

### Correlation Matrix

Using the XLSTAT add-on statistical software for Excel, Correlational analyses were used to examine the relationship between 32 different indicators (including VRE installed capacity share) in the 131 countries in the data set. The correlation matrix explores relevant unexpected results that can be later examined in advanced research.

Results indicated a low inverse relationship between hydro as a share of the total, and wind and solar installed capacity *r(131)= -.04, p< .03* two-tailed. The results suggest that hydro as a share of total correlations to all other indicators were negative, but only four governmental indicators were statistically significant and were lower or equal to *r(131) = -.06, p< .05,* two-tailed. This result suggests that hydropower systems have a weak correlation to the examined indicators.

According to the World energy council (2019), Hydropower is the most reliable and flexible of RE resources. Hydropower systems are capable of meeting base load requirements as well as peak and unexpected electricity demand. In 2016, hydropower was the leading renewable source for electricity generation globally, supplying 71% of all renewable electricity.

A possible explanation for the weak correlation founded is that both poor and reach countries such as Norway, Canada, Brazil, Albania, Mozambique, Zambia and more, with vast hydropower potential and more than 50, 70 or even 90% hydropower installed capacity have no need or motivation to invest heavily in renewable energy. A significant relationship was found between the World Bank governmental indicators as can be seen in the following table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Correlation matrix, coefficients of determination (Pearson): | | | | |  |  |
|  | wind and solar as share of total | Voice and account- ability | Political stability & absence of violence | Government effectiveness | Regulatory quality | Rule of law | Control of corruption |
| wind and solar as share of total |  | 0.299 | 0.131 | 0.263 | 0.266 | 0.280 | 0.268 |
| Voice and accountability | 0.299 |  | 0.566 | 0.579 | 0.634 | 0.687 | 0.648 |
| Political stability & absence of violence | 0.131 | 0.566 |  | 0.561 | 0.540 | 0.621 | 0.638 |
| Government effectiveness | 0.263 | 0.579 | 0.561 |  | 0.842 | 0.884 | 0.823 |
| Regulatory quality | 0.266 | 0.634 | 0.540 | 0.842 |  | 0.871 | 0.757 |
| Rule of law | 0.280 | 0.687 | 0.621 | 0.884 | 0.871 |  | 0.901 |
| Control of corruption | 0.268 | 0.648 | 0.638 | 0.823 | 0.757 | 0.901 |  |
|  | *p-values (Pearson) are different from 0 with a significance level alpha<0.001, 131 Observations* | | | | | |  |

Table 1 Relationship between World Bank governmental indicators

Finding this high correlation between the different indicators means the similarity between them will be relevant and should be considered in the following methods. The matrix found that 17 indicators out of the 31 examined, showed a significance level of correlation (P<0.05), to share of installed VRE as can be seen in Table 2.

|  |  |
| --- | --- |
| **Correlation to VRE as a share of total** | 131 Observations |
| **Indicators** | **correlation coefficient** |
| hydro as share of total *(Negative correlation)* | 0.037\* |
| GDP per capita | 0.219\*\*\* |
| Democracy | 0.139\*\*\* |
| Feed-in tariff | 0.079\*\* |
| Net metering/net billing | 0.048\* |
| Transport obligation/mandate | 0.199\*\*\* |
| Heat obligation/mandate | 0.155\*\*\* |
| Tradable REC | 0.177\*\*\* |
| Capital subsidy, grant, or rebate | 0.068\*\* |
| Investment or production tax credits | 0.089\*\* |
| Public investment, loans, or grants | 0.045\* |
| Voice and accountability | 0.299\*\*\* |
| Political stability and absence of violence | 0.131\*\*\* |
| Government effectiveness | 0.263\*\*\* |
| Regulatory quality | 0.266\*\*\* |
| Rule of law | 0.280\*\*\* |
| Control of corruption | 0.268\*\*\* |
| *Note: \* P < 0.05, \*\* P<0.01, \*\*\* P<0.001* |  |

Table 2 Variables correlation to VRE

The correlation results show that direct (ex. feed-in tariff) and indirect (ex. rule of law) variables affect the VRE penetration rate. It can be seen that the variable with the strongest correlation is Voice and Accountability (*r(131)= .299, p< .0001*), which relate to citizens ability to play a part in selecting the government, freedom of expression, association, and a free media. The last five indirect indicators in Table 2 that are highly relevant (*r(131)> .25, p< .0001*), are related to the quality and stability of the government. Direct regulation that supports renewable energies such as feed-in tariff, net metering has a much weaker link.

The world bank RISE indicators were analyzed separately due the comparatively small (105) amount of countries in the RISE database as in Table 3. The multiple regression for all these indicators shows high correlation to VRE *r(105)= .363, p< .0001.* These indicators are not part of the later analysis due to the similarity between the RISE indicators to the ones taken from other sources, and the different countries observed.

|  |  |
| --- | --- |
| **RISE Correlation to VRE as a share of total** | **105 Observations** |
| **Indicators** | **correlation coefficient** |
| Overall-Score Renewable Energy | 0.360\*\*\* |
| Legal framework for renewable energy | 0.038\* |
| Planning for renewable energy | 0.248\*\*\* |
| Incentives and regulatory support for renewable energy | 0.296\*\*\* |
| Attributs of financial and regulatory incentives | 0.088\*\*\* |
| Network connection and use | 0.235\*\*\* |
| Counterparty risk | 0.113\*\*\* |
| Carbon pricing and Monitoring | 0.264\*\*\* |
| *Note: \* P < 0.05, \*\* P<0.01, \*\*\* P<0.001* |  |

Table 3 RISE indicators correlation to VRE

### Multiple Regression

Part of the paper with Stefan, will be added later…

### SVM analysis

Support Vector Machine (SVM), as described in detail in the Methodology chapter, is an analytical tool that analyzes a large number of variables correlation to the discrete dependent variable. This method is prevalent in medical research but not in the fields of social science, governmental studies, and public policy. SVM is a steady algorithm that can cope with large sets of predictors at once. This method may be particularly useful in our case, in which we have 18 indicators with significant correlation to VRE penetration.

As far as I know, based on in-depth research in search engines as of Google Scholar and Microsoft Academic, no previous research has investigated these issues with the following methods. Only a few studies in the field of public policy, energy policy, and renewable energy that uses SVM. Most of the articles related to renewable energy and using advanced methods of machine learning and SVM focus in the fields of radiation and wind power, network effects, consumption forecasting, and future production capacity. Using this advanced method, which enables more accurate forecasting of various indicators, including policy and regulatory frameworks with approximately 80% accuracy for the introduction of renewable energies based on a small dataset (131).

SVM is an excellent analysis tool for discrete variables. In order to construct a matrix in which the dependent variable is discrete, we divided the 131 countries of the world into four groups according to the percentage of penetration of renewable energies.  
Group 1 – 0-1% - 72 Countries in which the VRE installed capacity is up to 1% of the total installed capacity. This group embraces mainly emerging markets such as Russia, African countries, East Asia, South America, and the Middle East.

Group 2 – 1-10% - 31 Countries in which the VRE installed capacity is more than 1% of the total installed capacity, but less than 10%. This group includes the United States, Canada, Brazil, India, Switzerland, Finland, and other developed and developing countries.

Group 3 – 10-20% - 17 Countries in which the VRE installed capacity is 10-20% of the total production capacity. This group is smaller and includes countries that have promoted the field in practice over the last decade and have attained relatively high production capacity. Countries such as France, Sweden, Estonia, and African countries such as Ethiopia and Nicaragua, where overall production capacity is low but a high percentage of solar energy thanks to many UN factories and investments (add quote)  
Group 4 – 20-50% - 11 Countries in which the VRE installed capacity is more than 20% of the total production capacity. This small group includes the pioneer countries that led the world's renewable energy market. The group consists mostly of European countries - Germany, Denmark, Belgium, Spain, Portugal, Britain, Italy, Ireland and Cyprus and two representatives from South America - Uruguay and Honduras.

### results

XLSTAT software (version 2018.5) was conducted for the analytic part. SVM was configured in linear and separable form, Multiclass with One versus all, and rescaling options. In these options, the algorithm selects a hyperplane that separates between the set of observations into two distinct classes. The algorithm maximizes the distance between the hyperplane and the closest observation of the training set. One binary model per class was created, where the dependent class is reserved, and all the other classes are merged into one class, and quantitative explanatory variables are rescaled between 0 and 1.

The 17 indicators with correlation to VRE (Table 2) for all 131 countries were used in order to train the SVM classifier. In order to check the classifier, three independent testings were made. The classifier was trained on 129 or 128 countries, and 2 or 3 randomly countries were chosen as a test set. The results of the SVM classifier were excellent, these results go beyond previous methods, reaching 80.15% persistence on the total, as shown in Table 4. The SVM classifier can predict 94.44% accuracy for countries in group 1, 61.29% for group 2, 58.82% for group 3, and 72.73% for countries in group 4.

Based on the results of the SVM algorithm, given a new country for which we have the independent variables selected, we can calculate the percentage of renewable energy installed in this country. Another ability derived from the use of SVM is the ability to perform simulations and thus to examine the variables to be treated. For example, a country with a known regime, population, territory, and given economy status, that wishes to increase to 15% of VRE can examine which mix of the many variables introduced here has the most substantial influence on the rate of penetration in this country.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SVM Confusion matrix for the training sample[[1]](#footnote-1) | | |  | |  |  |
| from\to VRE installed capacity | Total Countries | % correct |  |  |  |  |
| 0 – 1% | 72 | 94.44% |
| 1 – 10% | 31 | 61.29% |
| 10-20% | 17 | 58.82% |
| Above 20% | 11 | 72.73% |
| Total | 131 | 80.15% |

Table 4 SVM Results

This method attains a statistically significant improvement compared to previous methods. The results show that the SVM classifier is superior to multiple regression and other methods with the ability to reach high correction level with only moderate amounts of samples while avoiding common problems and limitations.

### CHAID Classificaion Trees

CHAID stands for Chi-squared Automatic Interaction Detector. It is one of the oldest tree classification methods originally proposed by Kass (1980). The goal of classification trees is to predict or explain responses on a categorical dependent variable. For regression-type problems (continuous dependent variable), other techniques (as of QUEST algorithm) are not applicable, so only CHAID and C&RT can be used. CHAID will build non-binary trees that tend to be "wider". This has made the CHAID method particularly popular in market research applications: CHAID often yields many terminal nodes connected to a single branch, which can be conveniently summarized in a simple two-way table with multiple categories for each variable or dimension of the table. (Hill, T. & Lewicki, 2007)

### Design of the classification\ Decision Trees

The main focus of this study was to analyze data classification and overview of the relationship between the policy instruments. A correlation matrix was performed with all data obtained in the previous section. The dataset includes indicators from four different groups: Policy instruments, governance indicators, development indicators, and fiscal & market indicators.

Since the plan is to use the CHAID tree results for the BNN section, the first step in handling the data is to reduce the indicators that are not controllable by the government and to use only handy policy instruments. As an example, coal and gas reserves are not controllable variables. Other indicators like democracy, Political stabilit, and absence of violence, Control of corruption, and others were examined but are not part of the final tree since the BBN tool goal is to build viable policy instrument. The recommendation to reduce violence, corruption or to change regime type is not practical ones for increasing VRE penetration. The following indicators entered to the CHAID tree: Feed-in tariff (FIT), Net metering/net billing (NET), Tradable REC (T-REC), Capital subsidy, grant, or rebate (CS), Investment or production tax credits (INV), Public investment, loans, or grants (PI), and Regulatory quality (RQ).

The CHAID decision tree models were designed from the learning set with the specific software of XLSTAT ver. 2019.1.2.56803 (N=131, Maximum tree depth- 3, Significance level (%)- 5, Merge threshold (%)- 5). I used the CHAID algorithm with a maximum tree depth of 3 sons to avoid obtaining a too complex tree. Many iterations were examined in order to find the best VRE installed capacity confusion matrix, in which the 131 countries are divided to three groups – 0 – No installed capacity of VRE, 1 – 0.001% - 6%, 2 - > 6%. Since 67 countries in this survey have zero VRE, in order to emphasize the weight of the different variables for the countries with VRE > 0%, another classification tree was tested with the same variables but with the 64 countries. The countries were divided into four groups – 0 – 0.01%-3% of installed capacity of VRE, 1 – 3% - 10%, 2 – 10% - 20%, 2 - > 20%.

### Results

This canonical representation of the data enabled us to make a distinction between the different categories and to get the best results of 75.57% well-classified observations on average. The confusion matrix shows a good estimation for no VRE (95.52) and more than 6% of VRE (75.0) but poor estimation for the low than 6% of VRE (20.83)

Confusion matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| from \ to |  | 0 | 1 | 2 | Total | % correct |
| 0 | No VRE | 64 | 0 | 3 | 67 | 95.522 |
| 1 | <6% VRE | 11 | 5 | 8 | 24 | 20.833 |
| 2 | >6% VRE | 10 | 0 | 30 | 40 | 75.000 |
| Total |  | 85 | 5 | 41 | 131 | 75.573 |

Table 5 CHAID well-classified observations

The model removed the less significant indicators of INV and CS and used only five indicators – FIT, NET, T-REC, PI, RQ.

Tree structure:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nodes** | **Objects** | **%** | **p-value** | **Purity** | **Split variable** | **Values** | **Parent node** | **Sons** | **Predicted values** |
| **# 1** | 131 | 100.0% | < 0.0001 | 51.1% |  |  |  | 2; 3 | 0 |
| **# 2** | 79 | 60.3% | 0.001 | 73.4% | RQ | <= 63.46 | 1 | 4; 5 | 0 |
| **# 3** | 52 | 39.7% | 0.031 | 61.5% | RQ | > 63.46 | 1 | 6; 7 | 2 |
| **# 4** | 59 | 45.0% | 0.039 | 84.7% | NET | <= 0 | 2 | 8; 9 | 0 |
| **# 5** | 20 | 15.3% | 0.027 | 40.0% | NET | > 0 | 2 | 10; 11 | 1 |
| **# 6** | 31 | 23.7% | 0.027 | 48.4% | T-REC | <= 0 | 3 | 12; 13 | 2 |
| **# 7** | 21 | 16.0% | 0.149 | 81.0% | T-REC | > 0 | 3 |  | 2 |
| **# 8** | 29 | 22.1% |  | 96.6% | PI | <= 0 | 4 |  | 0 |
| **# 9** | 30 | 22.9% |  | 73.3% | PI | > 0 | 4 |  | 0 |
| **# 10** | 5 | 3.8% |  | 100.0% | RQ | <= 37.98 | 5 |  | 1 |
| **# 11** | 15 | 11.5% |  | 53.3% | RQ | > 37.98 | 5 |  | 0 |
| **# 12** | 11 | 8.4% |  | 54.5% | FIT | <= 0 | 6 |  | 0 |
| **# 13** | 20 | 15.3% |  | 65.0% | FIT | > 0 | 6 |  | 2 |

Table 6 131 countries Classification tree structure

Table 6 categorizes variables that differentiate the percentage of installed VRE in the 131 countries according to the five indicators. Figure 5, represent the same data in a graphical tree structure. The classification tree has 13 nodes. For each node, the table represents the number of countries at each node, the corresponding %, the purity that indicates the dominating category % of objects of the dependent predicted variable at this node, the parent and child nodes, the splitting variable and the intervals or values of the latter, and the value predicted by the node (Addinsoft, 2019).

For example, in this model node #1, the first branch has 51.1% with the predicted value 0. The tree separates the data into two groups (nodes 2,3) according to variable RQ segmentation value 63.46. Node #2 has 60.3% of the observations, and node #3 39.7%. Node #8 has 29 observations, 22.1% of all, with 96.6% purity for the dominant\predicted value 0. The different level of splitting allowed to separate the branches with good accuracy from the others and get a high level of purity in some of the nodes (100%, 96.6%, 84.7%, 81%). No branch was under 40% accuracy.

Feed-in tariff (FIT), Net metering/net billing (NET), Tradable REC (T-REC), Capital subsidy, grant, or rebate (CS), Investment or production tax credits (INV), Public investment, loans, or grants (PI), and Regulatory quality (RQ). I used the CHAID algorithm with a maximum tree depth of 3 sons to avoid obtaining a too complex tree. The model removed the less significant indicators of INV and CS

The decision tree represents how different variables affect the VRE penetration with different perceptions of policy instruments: indications of VRE regulation such as “Capital subsidy, grant, or rebate” and “Investment or production tax credits” are not important for classification. Although most of the variables that emerge as decision tree analysis were also identified as significant by the regressions and correlation matrix in the previous section, the classification tree clarifies their significance by detecting these conditional relations. The effects of “Regulatory quality” appears to be a dominant predictor while “Feed-in-tariff” effect only 23.7% of all.

131 countries Classification Tree

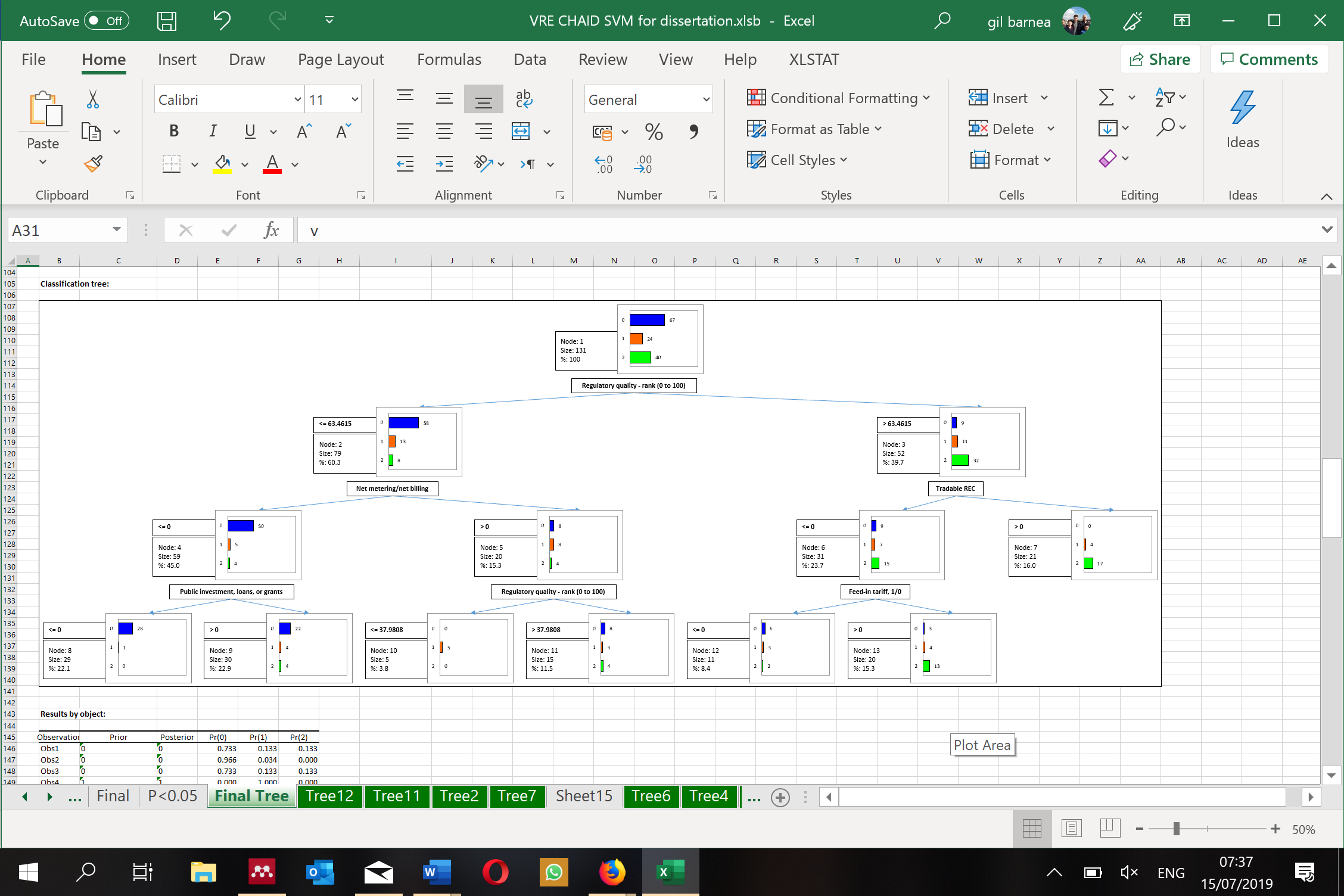


Figure 5 131 countries Classification Tree

The decision tree represents how different variables affect the VRE penetration with different perceptions of policy instruments: indications of VRE regulation such as “Capital subsidy, grant, or rebate” and “Investment or production tax credits” are not significant for classification. Although most of the variables that emerge as decision tree analysis were also identified as significant by the regressions and correlation matrix in the previous section, the classification tree clarifies their significance by detecting these conditional relations. The effects of “Regulatory quality” appears to be a dominant predictor while “Feed-in-tariff” effect only 23.7% of all.

The second classification tree was tested with the same variables but with the 64 countries. The confusion matrix for this tree is 76.9, 39.1, 76.4, 0 respectively with 50% well-classified observations on average. This model had a problem to isolate the 11 countries with a high level of VRE (>20%). This model (as can be seen in Table 7 64 countries Classification tree structure and Figure 6) highlight the effects of “Regulatory quality” along with “Net metering/ Net billing” and “Investment or production tax credit” as the dominant predictors for the 64 countries with VRE.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nodes** | **Objects** | **%** | **p-value** | **Purity** | **Split variable** | **Values** | **Parent node** | **Sons** | **Predicted values** |
| **# 1** | 64 | 100.00% | 0.007 | 35.94% |  |  |  | 2; 3 | 1 |
| **# 2** | 6 | 9.38% | 0.121 | 83.33% | RQ | <= 28.8462 | 1 |  | 0 |
| **# 3** | 58 | 90.63% | 0.049 | 39.66% | RQ | > 28.8462 | 1 | 4; 5 | 1 |
| **# 4** | 29 | 45.31% | 0.089 | 44.83% | NET | <= 0 | 3 |  | 2 |
| **# 5** | 29 | 45.31% | 0.029 | 44.83% | NET | > 0 | 3 | 6; 7 | 1 |
| **# 6** | 13 | 20.31% |  | 38.46% | INV | <= 0 | 5 |  | 0 |
| **# 7** | 16 | 25.00% |  | 56.25% | INV | > 0 | 5 |  | 1 |

Table 7 64 countries Classification tree structure

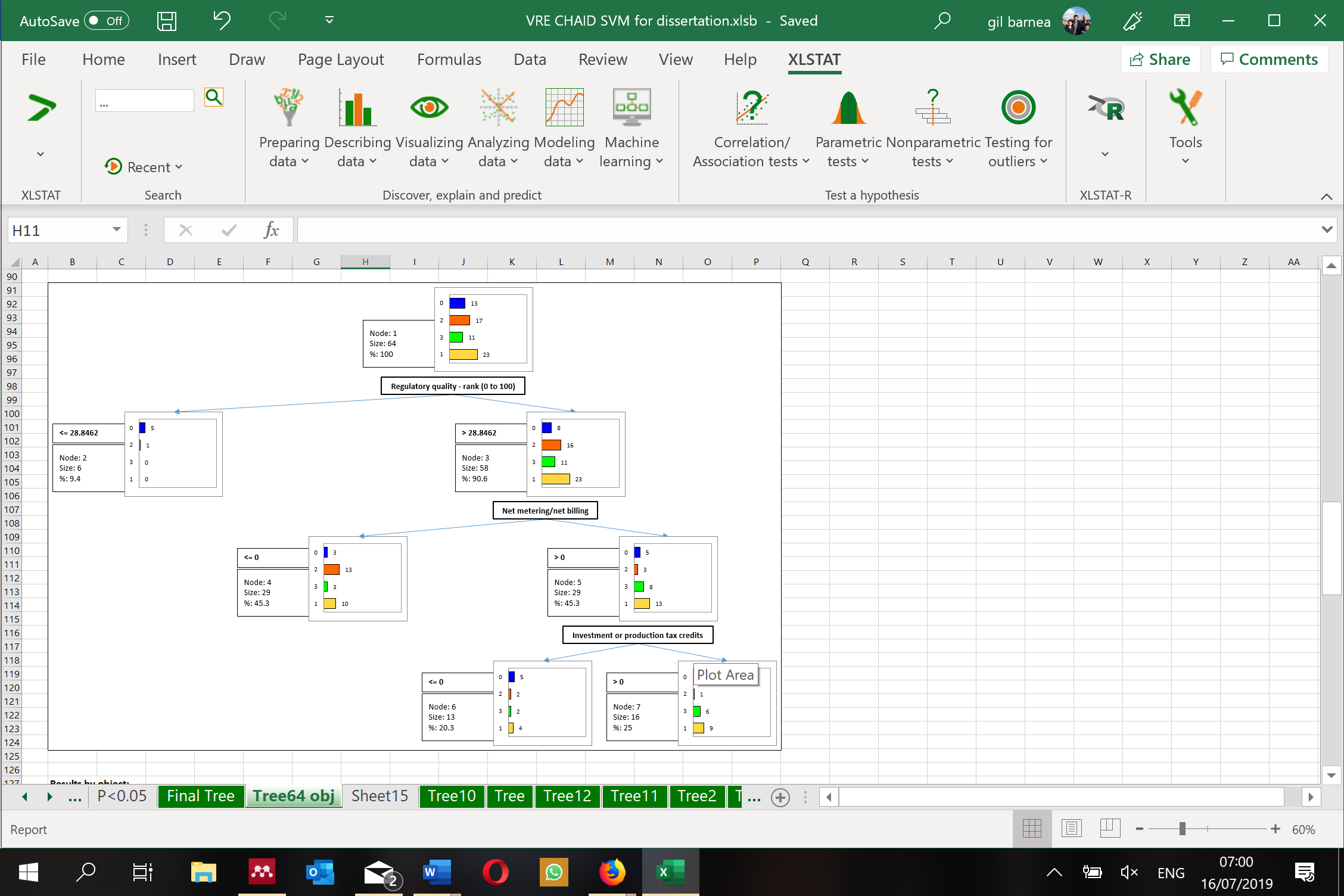


Figure 6 64 countries Classification Tree

### Disscussion

These CHAID classification tree results highlight different aspects of the indicators effect and the relationship between them. Regulatory quality that showed relatively high significance (*r(131)= .266, p< .0001*) in the correlation matrix. It should be noted that playing with the various parameters (e.g., levels of penetration or adding statistically significant indicators such as Government effectiveness, Rule of law, Voice and accountability) changes the tree generated and creates a different set of weights. It should be noted that the Indicators related to the quality of administration are constantly at the top of the tree and have a major influence on the different models examined. The use of classification tree arises dissimilar outcomes and highlight diverse connotation and signification from correlation matrix and multivariance regression equation. This model examined seven indicators that can be controlled by the government in order to connect it at the next step to the BBN. Building a classification tree based on all or other the significant indicators found by the correlation matrix, floated other indicators, and created a tree with other weights. These extra trees examined in the dissertation are not part of this paper since they are not relevant to the next step of BBN.

### BAYESIAN Model

### Bayes’ Theorem

Bayes theorem or rule is a part of the probability theory in statistic, used for calculating the probability of two or more independent events (Bayes & Price, 1763). The rule describes the probability of an event, based on prior knowledge of conditions that might be related to the event (Hooper, 2013; Stigler, 1983).

**Bayes Rule:**

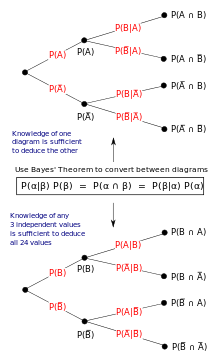


Figure 8 Bayes' Theorem sample (Wikipedia 2011)

For any two events, A and B:

**p(B|A) x p(A) = p(A|B) x p(B)**

p(A) - the probability of A

p(B) - the probability of B

p(A|B) - the probability of A given that B has occurred

p(B|A) - the probability of B given that A has occurred

Bayes' theorem depends on the interpretation of probability ascribed to the events. It defines the probability of an event as the limit of its relative frequency in a large number of trials. For example, as in Figure 8, an experiment is performed many times. P(A) is the share of outcomes with property A, and P(B) is of property B. P(B | A) is the share of outcomes with property B out of outcomes with property A, and P(A | B) the share of those with A out of those with B. Bayes' theorem is best visualized with diagrams, as shown in Figure 8. The two diagrams partition the experiment same outcomes by A and B in opposite orders, to obtain the inverse probabilities. Bayes' rule combines these different panels (WIKIPEDIA, 2011).

### Bayesian Belief Networks

Bayesian Belief Networks (BBN's) are networks of relationships. Named "Bayes" after Reverend Thomas Bayes, a British theologian, and mathematician who wrote down a basic law of probability around 1760, which is now called Bayes rule (Norsys, 2016, 2017).

BBN's are becoming a popular method and gaining prominence in scientific literature recent years in the fields of decision making in complex, data-poor situations (Borsuk, Stow, & Reckhow, 2004) as they can incorporate prior data from a diverse range of disciplines. BBN's are attractive decision support tool as they are well fitted to problems with small or incomplete datasets, and they are not restricted by minimum sample size (Kuhnert & Hayes, 2009).

In recent years, Bayesian belief network (BBN) models have been applied to numerous environmental and related applications (Krause et al., 2016).

A BBN is a graphical model representing the quantitative probabilistic relations between numerous events, which are related by their prior and conditional probabilities. The BBN is also a graphical structure used to describe the relationships between causes and effects in a diagram in which nodes represent events and probability description (Fenton & Neil, 2007). The network consists of the relationships among variables and nodes. The nodes in the network represent proportional variable of interests, arcs between nodes indicate the causal model influence or dependencies among the variables. Influence is propagated through a conditional probability table that specifies the influence between nodes. (Pearl & Russel, 2011).

Parameter estimation for a BBN model involves the determination of the conditional probability table for each node. This probability table may be approached using several qualitative or quantitative methods, including expert induction (Xian, 2012). In this work, the BBN could be used to analyze and predict the probability of penetration pace dependence on the results of the CHAID classification tree.

Several software applications are available for implementing BBNs (Small, 2016). I utilized Norsys Netica version 6.03 to develop decision nets, and influence diagrams model, in this study (Norsys, 2017). The software can create unidirectional causal relations and can find the best relations between the nodes to get the highest penetration percentage.

### Methodology

Bayes nets are used for risk management, modeling ecosystems, sensor fusion, prediction and more. The net represents causal chains, that may be defined as cause-effect relationships between parent and child nodes. It can supply evidence of past events, and then Bayes net can foresee what is the most likely future outcomes. Bayes nets strength is their robustness to missing data and the ability to make the best possible prediction with existing data.

In this work, the BBN is used to analyze the probability to VRE penetration pace dependence on policy and regulation instruments that effects the amount of installed capacity of VRE as described in previous sections. In order to design a practical model that can be used for future policy planning, only the regulation and policy tools were part of the CHAID tree that analyze the connections and influence of the different instruments to be used in the BBN. From the CHAID section, 5 regulation indicators out of 7 are used - Feed-in tariff (FIT), Net metering/net billing (NET), Tradable REC (T-REC), Public investment, loans, or grants (PI), and Regulatory quality (RQ). The model removed the less significant indicators of Capital subsidy, grant, or rebate (CS), and Investment or production tax credits (INV).

The first step was to build a realistic Bayes net with connected nodes representing the relevant indicators from the CHAID classification tree using Netica tools. The Netica software defines different types of nodes that create the net. For example, A discrete node is a node representing a categorical or discrete variable with a well-defined limited set of states or possible values. A *nature node* in a Bayes net represents some variable of interest or a variable that cannot be directly controlled by the decision maker. In case that a nature node is called *deterministic node* when it has a functional relationship with its parents, and *chance node*, whereas the relationship, is probabilistic. A *decision node* represents in the net a variable under the control of the decision maker.  When the net is solved, a decision rule is found for the node which optimizes the expected utility. A *utility node* in a Bayes net is a node whose expected value is to be maximized while looking for the best decision rule for each of the decision nodes.  For each child node, a Conditional Probability Table (CRT) is created in order to define the relations between the nodes. (Norsys, 2016).

### Data

In our model, we had to build a net with the indicators and weights we found in the CHAID decision tree section. The complexity now is to build the correct BBN net. The goal of the BBN net is not only to analyze privet cases or countries but to predict the influence of the different indicators on the penetration process worldwide.

The probability model for this network is:

*P(VRE, FIT, NET, PI, RQ, T-REC) =*

*P(FIT)\*P(T-REC)\*P(PI)\*P(RQ)\*P(RQ|NET)\*P(VRE|FIT, NET, PI, RQ, T-REC)*

The raw data starting point or baseline for the BBN model is as follow -

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Indicator | RQ\* | NET | T-REC | PI | FIT |
| Num Countries > 0 | 131 | 45 | 28 | 82 | 71 |
| Countries% > 0 (YES) | 39.7% High | 34% | 21% | 63% | 54% |

Table 8 Raw data starting point

\* *For RQ all countries are above 0. The average is 54% but only 39.7% are in the “HIGH” branch.*

The CHAID tree weights of each indicator and influence on the BBN model–

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Indicator | RQ | NET | T-REC | PI | FIT |
| Countries | 131 | 79 | 52 | 59 | 31 |
| Weight | 37% | 22% | 15% | 17% | 9% |

Table 9 BBNs Indicators weight

The base line status of VRE penetration is 67 (51%) countries with 0% VRE, 24 (18%) with 0.1%-6% and 40 (31%) with more than 6%.

The BBN model as described in Figure 8 has 4 nature nodes with discrete values (FIT, T-REC, NET, PI) of Yes or No and one node with continues values (RQ) of 0-100%. The decision tree of VRE penetration and utility node for Prediction (used for the simulation phase). The connection between the nodes is based on the CHAID tree influence diagram. For example, RQ has influenced both on the NET indicator and VRE penetration. The schematic model based on the CHAID classification tree output can be seen in Figure 8.

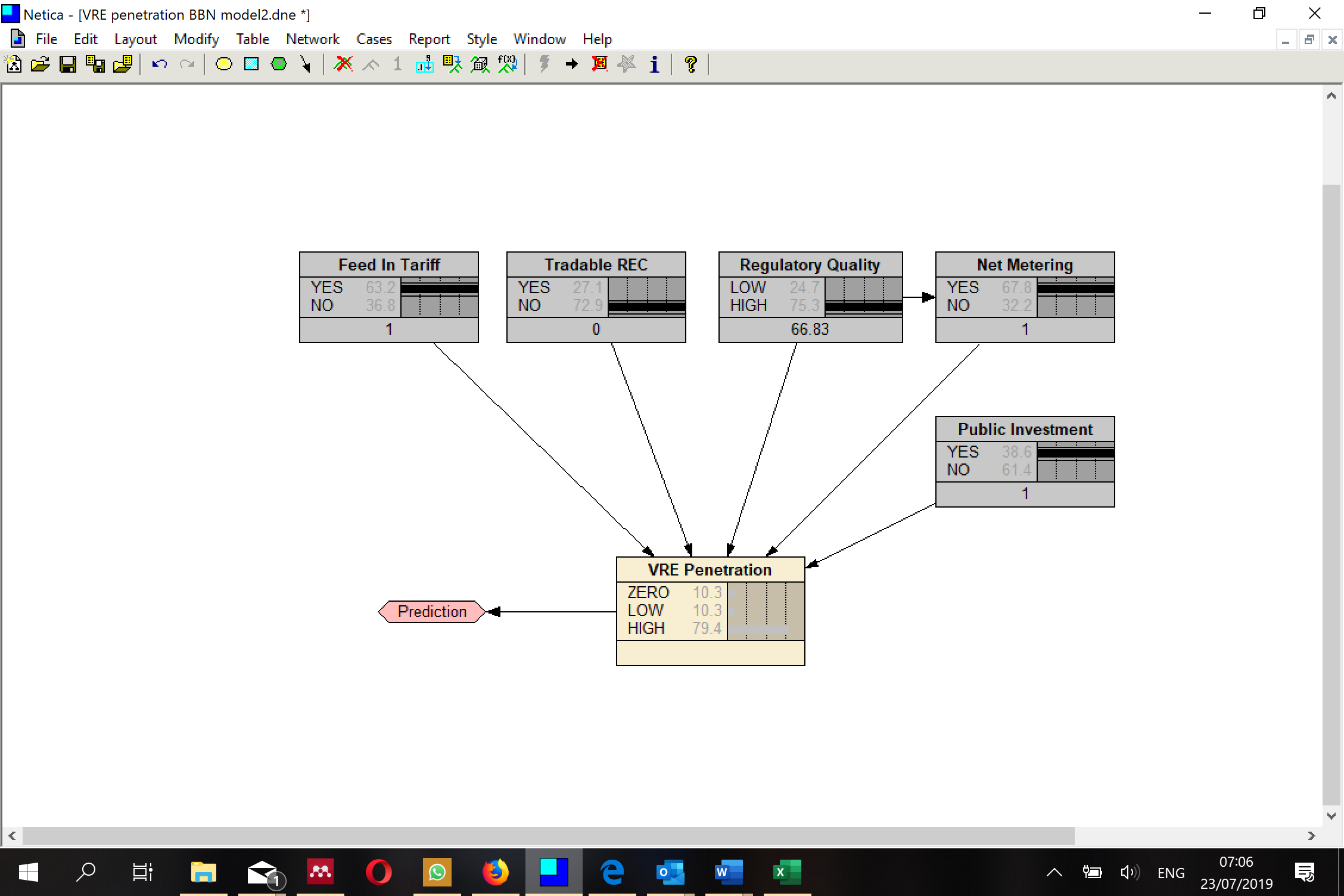


Figure 8 BBN model (The case of Greece as an example)

Figure 8 shows the penetration model as of Greece, and the effect of each variable on the total result. The set-up of Greece with 24.74% VRE installed capacity (HIGH) is FIT, NET and PI = 1 (Yes), T-REC = 0 (No) and RQ = 66.83%. For each child node, a conditional probability table is created in order to define the relations between the nodes. Figure 9 shows part of the CRT for the VRE penetration node.

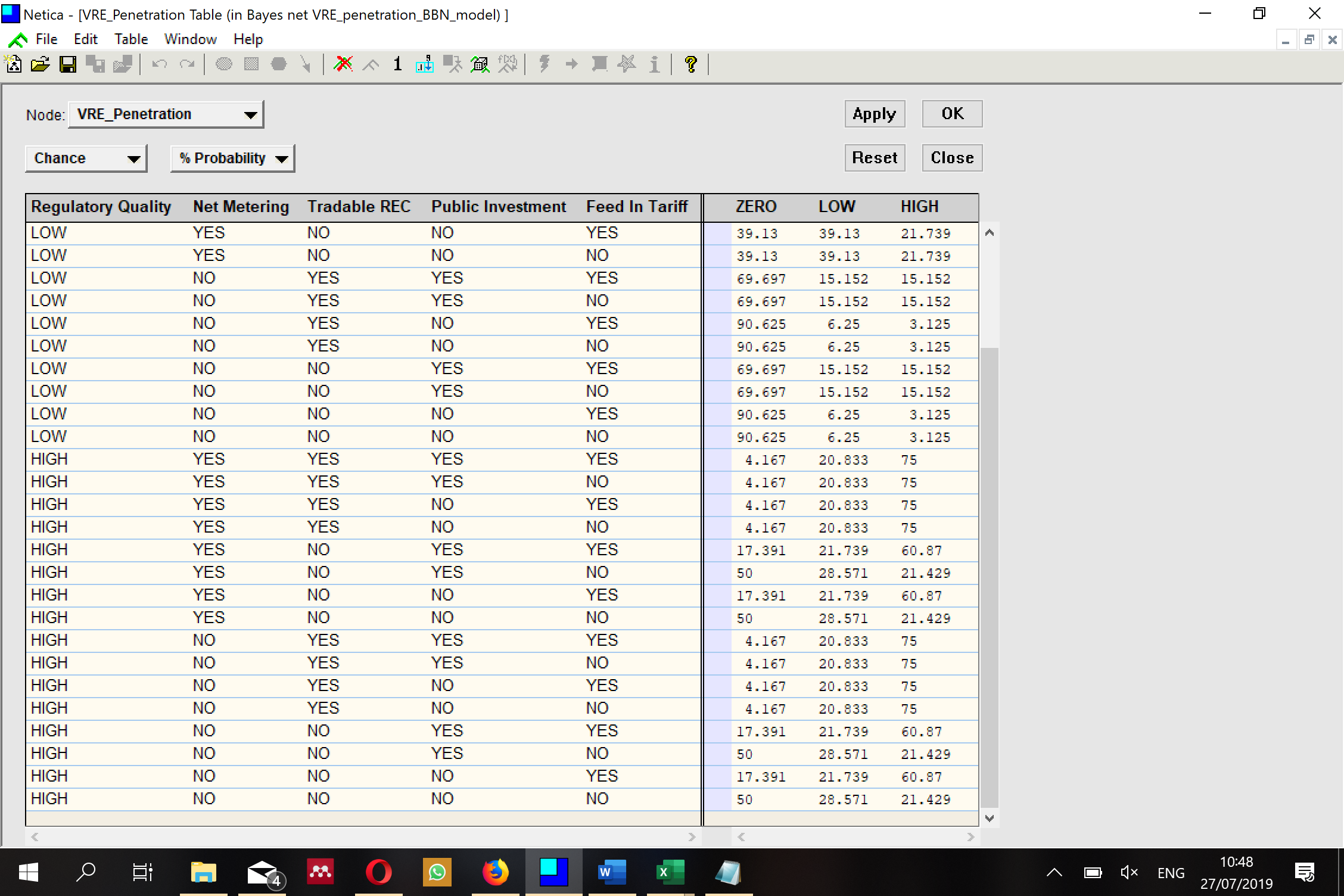


Figure 9 VRE penetration Conditional Probability Table (CRT) example

Using the Netica software package to develop the Bayesian belief networks, decision nets and influence diagrams model enables us to create many different occurrences and to see the effect of each one of them. The software creates unidirectional causal relations. We can ask the question – what would be the best scenario to increase penetration rate worldwide? The Netica can find the best relations between the nodes to get the highest penetration percentage.

### Results

The BBN model was examined with the different parameters of the baseline data of the countries in order to fit and adjust the model. Table 10 and the following simulation figures provide extensive results carried out with BBN that demonstrate the power of this technique that improves former methods results and obtain the most robust ones.

Table 10 summaries and illustrate some experimental results from the simulations made in this work. The baseline represents the starting data that was the input for the model. The calibration line demonstrates reverse testing – In case that all indicators are zero and RQ high for 42.7% of the countries – the model shows 75.8% of probability for zero VRE, 21.1% for low and 3.1% high. This is a logical and very reasonable analysis that confirms that the BBN can perform a reverse lookup for a theoretical scenario with no or very limited VRE penetration as a result of non-active policy instruments.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Penetration (%) | | | Indicators – YES (%) | | | | |
|  | **ZERO** | **LOW** | **HIGH** | **RQ High** | **NET** | **T-REC** | **PI** | **FIT** |
| Base line\ Reference | 51% | 18% | 31% | 39.7% | 34% | 21% | 63% | 54% |
| Calibration | 75.8 | 21.1 | 3.1 | 42.7 | 0 | 0 | 0 | 0 |
| Simulation D | 12.1 | 31.1 | 56.8 | 72.9 | 51.3 | 49.2 | 77.7 | 63 |
| Simulation E | 34.4 | 29.7 | 35.9 | 47.1 | 34.8 | 26.8 | 100 | 56.2 |
| Simulation F | 9.1 | 37.1 | 53.8 | 100 | 43.8 | 25.7 | 59.1 | 55.7 |
| Simulation G | 26.4 | 39.3 | 34.3 | 59.6 | 100 | 20.9 | 63.9 | 56.3 |
| Simulation H | 36.3 | 26.3 | 37.4 | 56.5 | 34.4 | 22.8 | 64.6 | 100 |

Table 10 Results table

Simulation D demonstrate a theoretical scenario of which the percentage of all indicators increase and then the penetration of VRE is growing significantly to 56.8% high, 31.1% low and the number of countries with zero penetration decline to only 12.1%.

Simulations E-H as can be seen in Figure 11 - Figure 17, are good scenarios to highlight the strength of the BBN. These set-ups are similar to the baseline except for one indicator that is changed to 100% yes. The results demonstrate two things.  First, increasing RQ (simulation F) will have a powerful effect on VRE penetration for countries without VRE and with a high amount of VRE. Second, the other three indicators – NET, PI, FIT (simulations E, G, H) will move countries from zero to the low amount of VRE but will have a minor effect on the high penetration rate of VRE.

This is an important finding in the understanding of the implications of each of the different variables and indicators on the deviation of VRE penetration. Moreover, our results demonstrate the importance of the different indicators in the different levels of VRE. In order to move from zero to the low amount of VRE, NET, PI, and FIT are good enough, but in order to increase penetration to more than 6%, RQ is essential.

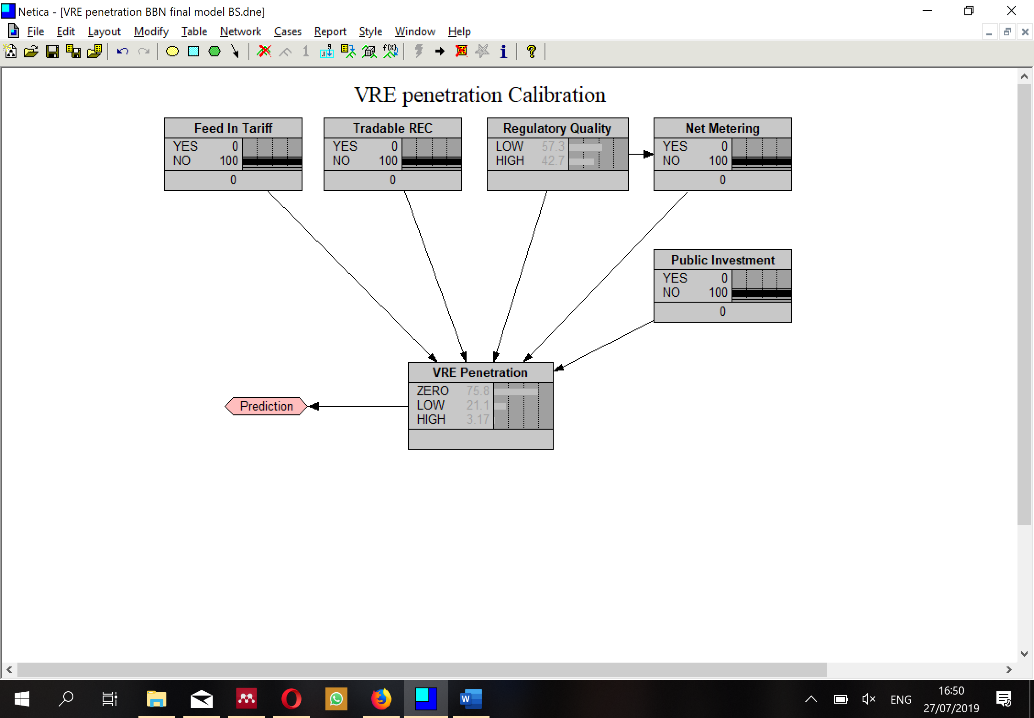
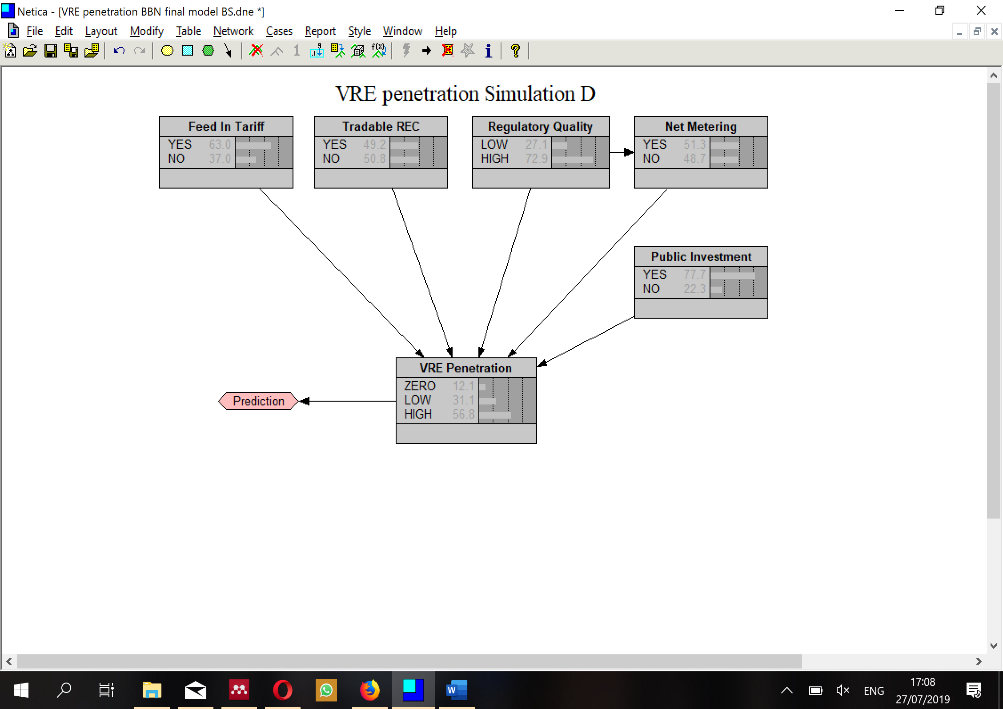
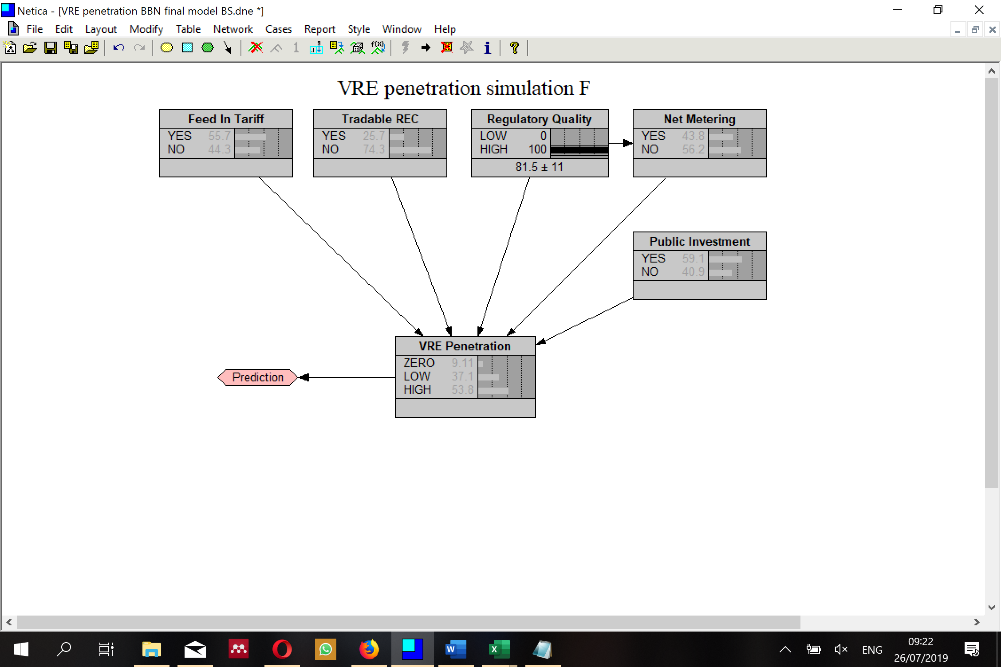
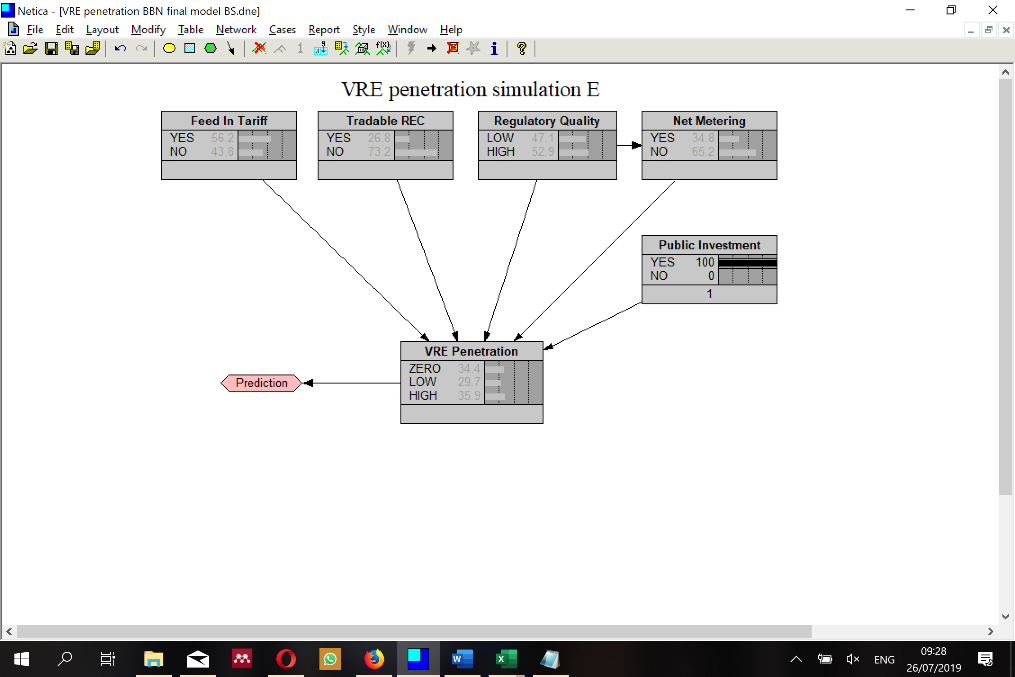


Figure 11 Simulation E

Figure 12 Simulation F

Figure 13 Simulation D

Figure 14 Calibration

Figure 15 Simulation E

#### 

Figure 16 Simulation H

Figure 17 Simulation G

### Discussion

The following table links the insights of the new methods with those of the traditional ones related to increasing penetration of VRE worldwide.

|  |  |
| --- | --- |
| Method | Output |
| Correlation Matrix | * Low inverse relationship between hydro and VRE * A significant relationship between the World Bank governmental indicators * 17 indicators out of the 31 examined showed a significant correlation to VRE * five indirect governance indicators are highly correlated. * World bank RISE indicators have similar correlation like REN21 indicators. * RISE overall score for VRE have 36% correlation. |
| Multiple Regression | * Part of the paper with Stefan, will be added later… |
| Support Vector Machine | * Predict 94.44% accuracy for countries with a very low amount of VRE, 61.29% for low, 58.82% for mid, and 72.73% for countries with high amount. * Can examine which mix of the variables introduced has the most substantial influence on the rate of penetration in a given country |
| CHAID Classification tree | * The effects of RQ appears to be a dominant predictor while the FIT effect is less significant * Indicators related to the quality of administration are consistently had a significant influence on the different models examined. |
| Bayesian Belief Network | * In order to move from zero to low amount of VRE, NET, PI, and FIT are proper policy tools, but in order to increase penetration, RQ is significant. * increasing RQ will have very strong effect for countries without VRE |

Table 11 Summery of outputs

These hands full insights we got, go beyond previous reports, showing that using advanced methods with more common and familiar ones can yield to significant understandings from a small dataset of observations and diverse indicators.

Overall, these good discoveries are in accordance with findings reported by world bank and REN21 related to policy instruments importance to increase VRE penetration pace. When comparing these results to those of other studies, it must be pointed out that this study validates the importance of indirect governance indicators like Voice and accountability, Political Stability, Absence of Violence, Government effectiveness, Regulatory quality and so.

The indicators with a strong impact in the correlation matrix and regression model also appear in the CHAID classification analysis as dominant indicators with very important relationships to penetration of VRE. Beyond those indicators, however, CHAID and BBN highlights different secondary variables like regulatory quality and public investments.

However, when comparing our results on a cluster level – policy instruments, governance indicators, development indicators, and fiscal & market indicators, the data reveals significant differences between the clusters. It must be pointed out that governance indicators (ex. Regulatory quality) happen to be the most influential parameters, and development ones, the less substantial.

The variety of outcomes demonstrate the dissimilar analysis offered by the different techniques. The differences are hardly surprising given the differences in these methods: the regression is a statistical technique based on least squares, while the later techniques are based on machine learning and Bayesian networks.

### Conclusion

The results of this chapter found clear support for the the main research question of this thesis. The present study confirm that governmental regulations and supporting policy plays a vital role in promoting the new market of VRE.

World-Bank 2018 RISE report highlight a significant improvement in VRE policies worldwide. the number of countries with advanced policy frameworks for sustainable energy has more than tripled over the past years (World-Bank, 2018b).

From the results above, it is clear that these steps are only part of the steps needed since indirect governance indicators are importance as much as direct policies and regulations.

The above findings show that for a given dataset, or a given theoretical model, a combination of statistical and machine learning techniques allows a broad and significant analysis of the probability and effect of different variables. This mixture brings significant conclusions that help investors, policy and decision makers a broader analysis of the meanings derived from the small number of influencing indicators. The limitations of the present work naturally include the small number of indicators and samples. Building a wider net with a larger dataset will enable a more precise model with higher accuracy in analysis and prediction. Another limitation in this work involves the issue of emerging new market with rapid changes in policies, decreasing price and improving technologies that yield to a complexity and less clear-cut data.

The results demonstrated in this chapter use state of the art methods. The BBN methodology has been demonstrated to be a versatile tool for making prospects for the future explicit, combining them with scenario analysis, and quantifying of different indicators. The CHAID decision tree and Bayesian networks are valuable tools for international organizations that investigate global trends and compare countries level due to the ability to work with small amount of observations.

### Summery

The analysis performed in this chapter at the global level gives an interesting picture of VRE's penetration processes depending on the many indicators examined. Of the 131 countries examined, more than half are at the beginning of the way and have only a fraction of a percentage installed capacity. 31 are at the beginning of the road with less than 10% renewable energy, 17 with significant production capacity between 10-20%, and 11 countries including the pioneer ones, of are leading the market forward. The comparison between hydroelectric and VRE shows a low negative correlation and shows that both developing and developed countries, with natural resources suitable for producing clean and cheaper electricity from hydro sources, will develop this field prior to VRE. Unexpectedly, there is a more significant correlation between the indirect political indicators like political stability, regulatory quality and rule of law as opposed to the direct policy instruments like feed in tariff, net metering and capital subsidy and the penetration rates of VRE. Regression analysis of multiple variables together shows that the regulation and policy instruments has a stronger effect than market instruments which can be explained in this stage as a consequence of premature growing market.

The combination between methods such as multiple regression, Correlation matrix, CHAID SVM, BBN used in this chapter, presented a very broad picture of connections, correlations, and networks of influence between the various indicators and the process of penetration of renewable energies.

The market creation theory on which this research is based, is highly connected to the analysis conducted in this chapter. It shows the pace of penetration process and the gap between the small number of pioneer countries that led the market forward during the last decade and most of the countries in the world that are far behind. This chapter clearly shows that the penetration process and creation of new market is now in the transition between the germination stage and the initial stabilization stage. The analysis found that pioneer countries are at the forefront are due to governmental indicators, policy issues, and regulatory factors that influence the creation of the new market that is currently under way.

Machine learning and artificial intelligence tools and methods that have evolved over the last decades using the rapidly increasing computational capabilities. These methods include vague logic, neuronal networks, influence networks, and Bayesian networks. These tools help to address a wide range of problems, but none of these methods provide a complete answer (Krigman, 2011). Although there is a congruence between the sets of solutions that each method can provide, each method is heading for at a certain type of problems. In recent years, many studies have been conducted to produce hybrid tools that combine two or more methods to solve indissoluble problems and overcome the disadvantages of each single method with the benefits of the other.

The use of several dissimilar methods, allows an in-depth and multi-dimensional data analysis on a given moderate data set. Each of the methods used, examines the data from different view and exposes other aspect of the relationship between the various indicators. The use of correlation matrix found strong connections between the political variables and the weak link to hydropower. The multiple regression indorsed for a more profound and meaningful analysis of the cumulative effect and the separation of market forces from regulatory ones. The SVM found a very high conjecture of 80.15% regarding the penetration of renewable energies. CHAID Decision Tree displays the assembly between the various variables graphically and emphasize the meaning of these connections. Analyzing the weights in the CHAID tree phase, enables building Bayesian Belief Network – BBN graphical model in which the indicators are the nodes represent events and probability description and CHAID tree weights creates the arcs between nodes and indicate the causal model influence.

### Bibliography

Addinsoft. (2019). XLSTAT statistical and data analysis solution. *Addinsoft*.

Antonio Erias, Cansu Karaka, Corinna Grajetzki, James Carton, Mekalia Paulos, Pirjo Jantunen, Prajwal Baral, Samal Bex, J. M. V. (2016). World Energy Resources 2016. *World Energy Council 2016*, 6–46. https://doi.org/http://www.worldenergy.org/wp-content/uploads/2013/09/Complete\_WER\_2013\_Survey.pdf

Antonio, J., Jennifer, C., & Vreeland, J. R. (2010). *Democracy and dictatorship revisited*. https://doi.org/10.1007/s11127-009-9491-2

Banerjee, u. G., Sinton, A., Moreno, J., Primiani, T., & Seong, J. (2016). *A Global Scorecard for Policy Makers*.

Bayes, T., & Price, R. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*.

Berwick, R. (1990). *An Idiot’s guide to Support vector machines (SVMs) R. Berwick, Village Idiot SVMs: A New Generation of Learning Algorithms*. Retrieved from http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf

Bock, T. (2019). What is a Correlation Matrix? | Displayr. Retrieved May 28, 2019, from https://www.displayr.com/what-is-a-correlation-matrix/

Borsuk, M. E., Stow, C. A., & Reckhow, K. H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, *173*(2–3), 219–239. https://doi.org/10.1016/J.ECOLMODEL.2003.08.020

Boser, B., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. *The Fifth Annual Workshop of Computational Learning Theory*, *5*, 142–155.

Breiman L., Friedman, J. H., Olshen, R., & Stone, C. J. (1984). *Classification and Regression Tree*. *Wadsworth & Brooks/Cole Advanced Books & Software, Pacific California*.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*(1), 113–126. https://doi.org/10.1016/S0893-6080(03)00169-2

Clark, P. (2007). of Research Questions. *Journal of MixedMethods Research*, *1*, 207–211.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*, 273–297.

Fenton, N., & Neil, M. (2007). Managing Risk in the Modern World Applications of Bayesian Networks A Knowledge Transfer Report from the London Mathematical Society and the Knowledge Transfer Network for Industrial Mathematics MANAGING RISK IN THE MODERN WORLD Applications of Bayesian N. *London Mathematical Society*, 5–6. Retrieved from http://www.agena.co.uk/resources/apps\_bayesian\_networks.pdf

Hales, D. (2018). *REN21 Renewables 2018 global status report*. Retrieved from http://www.ren21.net/wp-content/uploads/2018/06/17-8652\_GSR2018\_FullReport\_web\_final\_.pdf

Hill, T. & Lewicki, P. (2007). *STATISTICS: Methods and Applications.* *statsoft, Tulsa,*. Retrieved from http://www.statsoft.com/Textbook/Classification-Trees

Hooper, M. (2013). Richard Price, Bayes’ theorem, and God. *Significance*, *10*(1), 36–39. https://doi.org/10.1111/j.1740-9713.2013.00638.x

IEA Beta. (2019). International Energy Statistics. Retrieved April 26, 2019, from https://www.eia.gov/beta/international/data/browser/#/?pa=00000000000000000000040g2&c=00000002000g000000000800000000000000000000000002&ct=0&ug=8&tl\_id=2-A&vs=INTL.2-7-CHN-MK.A&cy=2014&vo=0&v=C&start=2014&end=2015

Jain Rashmi. (2017). Simple Tutorial on SVM and Parameter Tuning in Python and R | HackerEarth Blog. Retrieved December 1, 2018, from https://www.hackerearth.com/blog/machine-learning/simple-tutorial-svm-parameter-tuning-python-r/

Kass G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *20*, 119–127.

Krause, J., Small, M. J., Haas, A., & Jaeger, C. C. (2016). An expert-based bayesian assessment of 2030 German new vehicle CO2 emissions and related costs. *Transport Policy*, *52*, 197–208. https://doi.org/10.1016/j.tranpol.2016.08.005

Krigman, S. (2011). *A Comparative Review Of Methods That Combine complex methods of BBN & fusy logic*. *OPEN University*.

Kuhnert, P. M., & Hayes, K. R. (2009). How believable is your BBN? *Proceedings of the 18th World IMACS*, (July), 13–17.

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data and a proposal. *American Statistic Assosiation*, *58*, 415–434.

Norsys. (2016). Tutorial on Bayesian Networks with Netica. Retrieved from https://www.norsys.com/tutorials/netica/secA/tut\_A1.htm

Norsys. (2017). Netica. NORSYS. Retrieved from https://www.norsys.com/netica.html

Pearl, J., & Russel, S. (2011). BAYESIAN NETWORKS. *UCLA*.

Raghavendra. N, S., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, *19*, 372–386. https://doi.org/10.1016/J.ASOC.2014.02.002

Small, M. J. (2016). Handout: Intro to BBN. Small.

Stigler, S. M. (1983). Who Discovered Bayes’s Theorem? *The American Statistician*, *37*(4a), 290–296. https://doi.org/10.1080/00031305.1983.10483122

Struhl Steven. (2018). *Classification tree methods AID, CHAID and CART*. Retrieved from https://canvas.jmu.edu/courses/1520248/files/79047309/download?...frd...

Vapnik, V., & Chervonenkis, A. (1964). A note on one class of perceptrons — Kernel Machines. *Automation and Remote Control*, *25*. Retrieved from http://www.kernel-machines.org/publications/VapChe64

Vieira, B. H. (2017). classification - How does a Support Vector Machine (SVM) work? - Cross Validated. Retrieved December 1, 2018, from https://stats.stackexchange.com/questions/23391/how-does-a-support-vector-machine-svm-work

WIKIPEDIA. (2011). File:Bayes theorem simple example tree.svg - Wikimedia Commons. Retrieved July 19, 2019, from https://commons.wikimedia.org/wiki/File:Bayes\_theorem\_simple\_example\_tree.svg

World-Bank. (2018a). Policy Matters, REGULATORY INDICATORS FOR SUSTAINABLE ENERGY. Retrieved from http://documents.worldbank.org/curated/en/553071544206394642/pdf/132782-replacement-PUBLIC-RiseReport-HighRes.pdf

World-Bank. (2018b). *RISE 2018 EXECUTIVE SUMMARY*.

World Energy Council. (2019). Hydropower.

Worldbank Bank & NRGI. (2018). WGI 2018 Interactive. Retrieved from https://info.worldbank.org/governance/wgi/#home

Xian, S. (2012). Extreme Event Outcomes , Scientific Study , and Implications for Decision Maker Risk Perception and Mitigation, 1–39.

1. VRE CHAID SVM for dissertation.xls [↑](#footnote-ref-1)