

Article

# Is Investing in Companies Manufacturing Solar Components a Lucrative Business? A Decision Tree Based Analysis

Sebastian Klaudiusz Tomczak <sup>1,\*</sup> , Anna Skowrońska-Szmer <sup>1</sup> and Jan Jakub Szczygielski <sup>2,3</sup>

<sup>1</sup> Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland; anna.skowronska-szmer@pwr.edu.pl

<sup>2</sup> Department of Financial Management, University of Pretoria, Pretoria 0002, South Africa; kuba.szczygielski@northumbria.ac.uk

<sup>3</sup> Newcastle Business School, Northumbria University, Newcastle NE1 8ST, UK

\* Correspondence: sebastian.tomczak@pwr.edu.pl

Received: 28 November 2019; Accepted: 15 January 2020; Published: 20 January 2020



**Abstract:** In an era of increasing energy production from renewable sources, the demand for components for renewable energy systems has dramatically increased. Consequently, managers and investors are interested in knowing whether a company associated with the semiconductor and related device manufacturing sector, especially the photovoltaic (PV) systems manufacturers, is a money-making business. We apply a new approach that extends prior research by applying decision trees (DTs) to identify ratios (i.e., indicators), which discriminate between companies within the sector that do (designated as “green”) and do not (“red”) produce elements of PV systems. Our results indicate that on the basis of selected ratios, green companies can be distinguished from the red companies without an in-depth analysis of the product portfolio. We also find that green companies, especially operating in China are characterized by lower financial performance, thus providing a negative (and unexpected) answer to the question posed in the title.

**Keywords:** CHAID; CRT; QUEST models; manufacturing companies; renewable energy sources; financial ratios

---

## 1. Introduction

The majority of the global energy supply is generated by the burning of fossil fuels. Progress and global economic growth have been accompanied by an increase in the consumption of fossil fuels, resulting in climate change and global warming. Importantly, fossil fuels are distributed unequally across the global geography, creating challenges related to access in unstable and conflict zones. This has resulted in a shift towards the use of renewable energy sources (RES), with the intention of reducing dependency on fossil fuels and the challenges associated with the use of fossil fuels. Renewable energy is locally available, clean, sustainable and eco-friendly and therefore offers an attractive alternative to fossil fuels. Consequently, researchers and policy makers have shown an increased interest in greener sources of energy. According to data [1] available from the International Energy Agency (IEA), the share of electricity generation from renewable sources in overall energy production in 2018 has exceeded 25% and is constantly growing. This is evident in the increased production numbers of renewable energy over the three preceding decades (see Figure 1). During recent decades, the share of energy from wind and solar photovoltaic (PV) in total production of renewable energy has increased significantly.

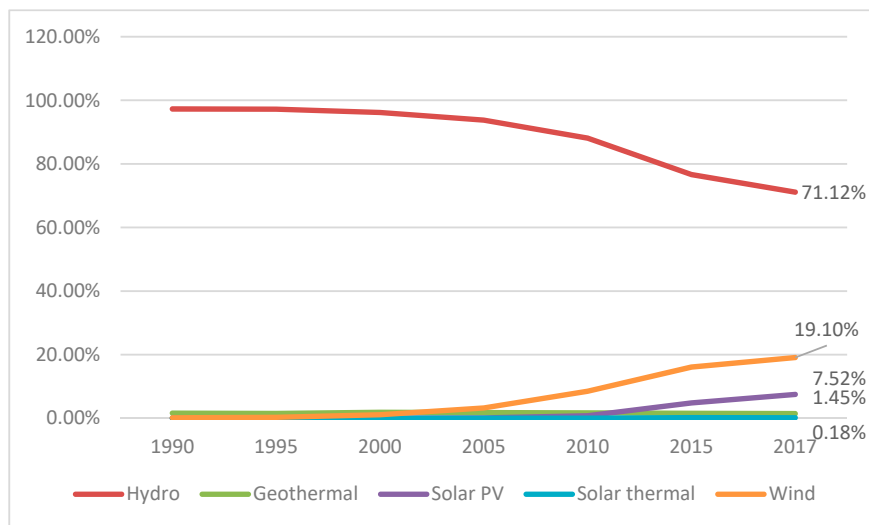


Figure 1. Global electricity generation from renewable sources, 1990–2017. Data source: [1].

Many countries have decided to promote renewable energy and to reduce dependence on fossil fuels and to do so, have established mandatory renewable energy targets. Such measures form part of respective governments’ legislated plans that require electricity retailers to source specific proportions of total electricity sales from RES within a fixed time frame. For example, the People’s Republic of China and India have a target level of 50% and 40% of total electric power generation from non-fossil energy sources, respectively. Renewable energies have been incorporated leading source of non-fossil energy. Specifically, the People’s Republic of China has established targets of 723 GW from RES by 2020. The underlying targets include increasing wind power capacity to 210 GW and solar energy to 110 GW [2]. In turn, the government of India set a target of achieving 175 GW from RES by 2022 which mainly consists of energy from solar (100 GW) and wind (60 GW) [3]. Moreover, Turkey, Thailand and Chinese Taipei have established RES targets as part of their total electric power generation. Turkey has set a target share of RES at 30% by 2030. Thailand has set a target level of 20% by 2036 and Chinese Taipei has a target of 8% by 2025 (see Figure 2). According to data provided by the IEA, none of the countries with set targets have achieved their targets, aside from Turkey which achieved its target in 2015.

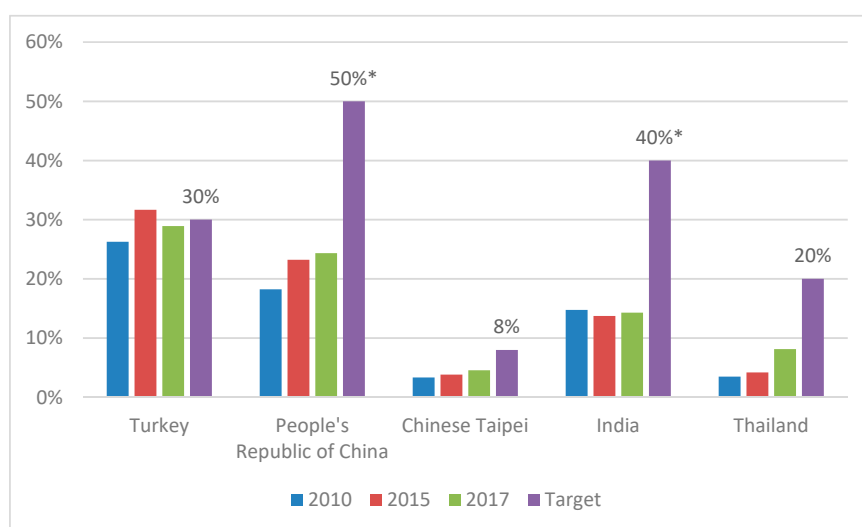


Figure 2. Electricity generation from renewable sources; share in the overall energy production for selected countries, 2010–2017 (in %). Data source—[1,4–6]. Target for Turkey, People’s Republic of China, India is for 2030, target for Chinese Taipei is for 2025 and target for Thailand is for 2036; \* means the percentage of non-fossil energy in total electric power generation.

Achieving these targets necessitates an increase in installed electric power generation capacity. Thus, the demand for manufactured solar modules, solar cells, solar silicon rods, solar wafers, solar power, solar photovoltaic products and equipment is increasing. This increase in demand is further compounded by the incentives offered by governments to promote investment in the PV industry and RES energy in general. For example, the Chinese government has set up national science and technology plans to support the research and development (R&D) into PV technology, the setting up of key laboratories that promote R&D in the industry and the promotion of demonstration projects in rural areas. Also, a feed-in tariff policy was implemented and exemptions from import tariffs were offered for imported equipment for domestic and foreign PV investment projects. The Chinese government also offered tax incentives for PV related enterprises in the form of R&D cost rebates, rebates for the purchases of special equipment and the amortization of costs for intangible assets and tax policies to promote PV power generation, high-tech enterprises in the PV sector and investment in R&D and PV production processes [7,8] provides an overview of the measures applied to promote investment in PV capacity within the EU-27 countries. The most popular measure to promote PV is the use of feed-in tariffs, which guarantee the price at which electricity is purchased. This is then followed by subsidies which are part of specific national programmes although subsidies are decreasing. Then, tax incentives are offered in almost half of the EU-27 countries. Broadly, these include tax credits, exemptions and reduced tax rates. For example, in Bulgaria, value added tax (VAT) for PV revenues is 10% as opposed to the normal rate of 15%. Tax deductions or tax credits are applied in Belgium, Denmark, France, Germany, Ireland and the Netherlands, reducing the amount of taxable income or the amount of tax due respectively. A number of countries have implemented tradable green certificates. For example, Belgium implemented a green certificate trading system for electricity, providing remuneration per certificate coming from a grid operator. Finally, soft loans have also been offered by a small number of EU countries. Solangi et al. [9] outline measures used to promote solar energy deployment in the US and Canada. In the US, incentives were formulated as early as 1978, with investment tax credits being offered for renewable energy technologies initially. Residential tax credits of 30% and business tax credits for investment in renewable energy, including solar energy, of 15% were offered respectively. Later on, residential and business investment tax credits were raised to 30% and extended, resulting in a doubling of installed PV capacity. In Canada, feed-in tariffs were implemented in 18 provinces, with the feed-in tariff implemented in Ontario credited with stimulating manufacturing and the growth of the Ontario PV industry. The Canadian government also provided loan guarantees, tax holidays and subsidies to support PV manufacturing. Sachu [10] outlines policies applied in the top 10 RES producing countries, including Germany. To promote the use of renewable energy, the German government mandated the purchase of renewable energy and offered large subsidies to renewable energy producers. A target was set to reach a specified level of solar PV capacity and feed-in tariffs were established. Sachu [10] notes that the policies applied in Germany have had a significant impact on reducing soft costs associated with solar installation, such as permitting, inspections, interconnection, financing and customer acquisition.

All these policies, which include targets and incentives, require that investment in RES companies increases. This can readily take place happen if RES related companies, specifically those that produce photovoltaic components, offer an attractive alternative to other companies that belong to the semiconductor and related device manufacturing sector and are lucrative overall. We contribute to the literature on RES by focusing on a class of companies that operate within the renewable energy value chain. These are companies that manufacture semiconductors and related solid state devices. We divide these companies into two groups. The green group comprises companies that manufacture solar modules, solar cells, solar silicon rods, solar wafers, solar power, solar photovoltaic products and related equipment for RES companies. This group is the focus of this paper. The red group comprises the remainder of companies that are not associated with RES companies).

We seek to identify ratios (i.e., indicators), which discriminate between companies within the manufacturing sector that produce PV components and those that do not. Once we have identified

such ratios, we can analyze these ratios to answer the question as to whether these companies show strong financial performance, that is, offer an attractive investment alternative. In the literature, there is great interest in evaluating returns for the energy sector, especially for companies that produce renewable energy. Relationships between returns on renewable energy stocks, changes in the oil price, equity indices, carbon prices and carbon pass-through rates, for example, Trück and Weron [11] examine convenience yields and risk premiums in the EU-wide CO<sub>2</sub> emissions trading scheme during the first Kyoto commitment period. They report that during their sample period, the EUA market has shifted from a period of backwardation to a period of contango with negative convenience yields. Bohl et al. [12] investigate the common factors driving the performance of German renewable energy stocks. They find that renewable energy stocks mirror ambiguity related to the future economic outlook faced by the industry. Henriques and Sadorsky [13] investigate the relationship between alternative energy stock prices, technology stock prices, oil prices and interest rates. Stock prices for alternative energy companies are found to be impacted by shocks to technology stock prices but shocks to oil prices do not have an impact on alternative energy companies. Inchauspe et al. [14] investigate the impact of energy prices and stock market indices on private investments in renewable energy, finding that these have shown substantial growth attributable to government policies, oil prices and evolving market liquidity. Kumar et al. [15] analyze the relationship between oil prices and alternate energy prices, finding that oil prices, stock prices for high technology firms and interest rates impact clean stocks. Managi and Okimoto [16] analyze the relationships between oil prices, clean energy stocks prices and technology stock prices. The impact of legislation on electricity future contracts is examined by Reference [17]. Studies also investigate the financial performance of firms operating in the energy sector [18–22]. For example, Capece et al. [18] focus on changes in the performance of the natural gas retail markets and analyze the financial statements of 105 companies by performing cluster analysis. They report that most companies in the sector perform well, with best performers belonging to existing business groups. Capece et al. [19] analyze the combined effect of regulatory measures and of the economic crisis on the performance of Italian gas companies. They observe that average measures of profitability rose until 2009 and then declined in 2009. This is attributed to the financial crisis and regulatory policies.

A number of studies, similarly to ours, evaluate the financial performance of firms that operate within the renewable energy sector using financial ratios. For example, Halkos and Tzeremes [20] evaluate the financial performance of renewable energy firms. They apply a data envelopment analysis and construct financial ratios to assess performance. Performance is found to be positively related to high levels of returns on assets and equity and by lower levels of debt. They also undertake a within sector analysis, finding that firms producing wind energy outperform firms generating hydropower. The dataset is limited to the Greek renewable energy sector and considers a limited number of ratios in the analysis of RES firm performance. Paun [21] analyzes the sustainability of the renewable energy sector in Romania. Their sample encompasses 91 major energy producers for the years 2012 to 2015. To analyze performance, they consider profitability ratios, measures of return on equity and measures of return on assets but do not consider key measures in the form of return on investment and the current ratio, given limitations in the data. They find that companies that produce fossil fuels perform better relatively to companies that produce green energy. They also report deteriorating performance, after 2013 and relatively low returns on equity for RES firms after 2014, suggesting that RES firms underperform other sectors. Also, returns on assets are found to decrease after 2013, this being attributed to changes in government and delays in issuing green certificates, leading to decreases in investment. Based upon ratio analysis, Paun [21] concludes that RES companies are close to financial distress. Tomczak [22] investigates whether powerplants that produce electricity using renewable energy sources are in a better financial position than those that rely upon traditional (fossil fuel) energy sources using a sample of companies in Baltic and Central European countries, comprising a total sample of 37 companies. In order to assess financial performance, he considers a total of 16 ratios indicative of four aspects, namely liquidity, profitability, turnover and debt—ratios

that are used in bankruptcy prediction. An analysis of ratios shows RES companies have lower return on assets and return on equity ratios than fossil fuel energy producers, which translates into lower interest from potential investors. Also, fossil fuel companies were found to be characterized by higher profitability but lower turnover ratios. Tomczak [22] concludes that investing in RES companies is not profitable business. Our study is similar to these studies; we also consider financial ratios, through the use of DTs, we identify and then analyze financial ratios to gain an insight into the performance of green companies.

Another strand of literature considers the relationship between corporate environmental performance and financial performance, with performance again measured by ratios, demonstrating the value of ratio analysis [23–26]. Clarkson et al. [23] investigate the factors that impact firms' decisions to engage in a proactive environmental strategy. The study is carried out using a sample of firms belonging to the four most polluting firms in the US, namely the pulp and paper, chemical, oil and gas and metals and mining industries. They consider how changes in environmental strategy impact profitability, liquidity and leverage. Analysis shows that becoming green leads to improved firm performance and that improved firm performance subsequently improves relative environmental performance. This is reflected by increasing return on assets, rising cashflows (profitability) and a decreases in leverage. Ruggiero and Lehkonen [24] show that there is a negative relationship between renewable energy production by electricity producers and short- and long-term financial performance. To measure financial performance, they use the return on equity and the return on assets and a firm's market value relative to total assets and regress these onto the volume of renewable energy produced, as a main variable of interest. Granger-causality is also tested. They find a negative relationship between measures of financial performance and the amount of renewable energy produced, attributing this to higher capital costs. Sueyoshi and Goto [25] examine the impact of environmental regulations in the US. Similarly to Ruggiero and Lehkonen [24], they use the return on assets as a measure of financial performance. They regress the return on assets for a 167 US electric utilities onto measures of environmental protection. Results indicate that environmental protection expenditure by US electric utilities has resulted in decreased financial performance. Sueyoshi and Goto [25] conclude that both the positive and negative aspects of environmental policies should be considered, given that emphasis is usually placed upon the positive aspects and negative aspects. Telle [26] challenges the methods applied in previous studies to conclude that firms that go green benefit financially. They note that one of the issues in studies that seek to relate financial performance to environmental performance is the presence of omitted variables. The omission of variables may result in an (erroneous) positive relationship between financial performance and environmental performance. Telle [26] shows that when the return on sales for Norwegian plants are regressed onto observable firm characteristics and omitted unobserved variables are controlled for, the positive impact of good environmental performance disappears. The study's significance lies in challenging findings of a positive impact of environmental performance on financial performance. The question of going green—and analogously—of producing components for the RES industry remains unanswered. Within the context of our present study, such findings suggest that green firms—and those related to RES—may not be characterized by improved financial performance relative to firms that are not green.

Finally, a number studies consider further aspects of the financial performance of firms that operate within the energy sector. For example, Pätäri et al. [27] investigate whether corporate social responsibility investments have an impact on corporate performance within the energy industry, finding that different aspects of corporate social responsibility impact either both profitability and/or market value. Arslan-Ayaydin and Thewissen [28] compare the financial performance of energy sector firms with different environmental scores, finding that those with good scores outperform financially those with poor scores. Sueyoshi and Goto [29] investigate the efficiency of national oil companies and aim to establish whether companies that are under public ownership outperform those under international ownership, showing that companies under national ownership outperform international oil companies under international private ownership in terms of efficiency.



None of these studies seek to investigate whether investing in companies manufacturing solar components is a lucrative business by classifying companies on the basis of ratios and then analyzing these ratios. To answer to this question, we apply a decision tree based analysis. We frame our analysis within the theoretical basis provided by literature on corporate failure and financial distress, which relies upon the classification of firms, as we do in this study [30,31]. The genesis of bankruptcy prediction models, is attributed to Beaver (1966) who uses 30 grouped ratios and information for failed and non-failed industrial firms to identify five ratios predictive of bankruptcy. His noteworthy conclusion is that ratios for distressed firms differ from those of healthy firms, making it possible to predict financial distress by discriminating between healthy and distressed firms. Altman [32] builds upon Beaver's [33] work and applies multiple discriminant analysis (MDA) to estimate a five-factor model predictive of bankruptcy for a sample of manufacturing firms. The joint consideration of ratios in Altman (1968) removes the ambiguity and confusion associated with interpreting individual ratios. Ohlson [34] proposes the use of logit analysis (logistic regression) to estimate an O-score model that combines financial ratios and indicators, yielding a probability of bankruptcy bounded between 0 and 1. The model provides a more precise and readily understood indication of the likelihood of bankruptcy and permits a greater range of outcomes than Altman's (1968) model. Frydman, Altman and Kao [35] propose a non-parametric solution in the form of decision trees (DTs) to classify distressed and non-distressed firms. DTs require no distributional assumptions or transformed variables. They can handle missing values and qualitative data, are readily understood and can incorporate misclassification costs. They apply these to manufacturing and retailing companies, finding that DTs are more accurate and minimize misclassification costs. Koh and Low [36] show that DTs outperform logit analysis and artificial neural networks in predicting bankruptcies. Similarly, Li, Sun and Wu [37] assess the performance of algorithmic implementations of decision trees for companies on the Shenzhen and Shanghai Stock Exchange, finding that these outperform other classifications methods—notably methods used in bankruptcy prediction. Gepp, Kumar and Bhattacharya [38] take a similar approach, showing that DT implementations are superior classifiers and predictors of bankruptcy for retailing and manufacturing companies. Fedorova, Gilenko and Dovzhenko [39] extend the application of bankruptcy prediction models to the Russian manufacturing sector by combining statistical methods with artificial intelligence techniques.

We build upon bankruptcy prediction literature that applies DTs to classify bankrupt firms and to predict bankruptcy. We apply DTs in a novel manner. While we do not aim to predict bankruptcy on the basis of financial ratios, we set out to determine whether companies that manufacture solar modules, solar cells, solar silicon rods, solar wafers, solar power, solar photovoltaic products and related equipment (green companies) can be differentiated from other enterprises in the sector that are not associated with RES companies (red companies) on the basis of financial ratios and whether these companies are in a better financial state. Based upon our analysis, the most critical ratios can be identified from a total of 62 ratios. It is hoped that using these ratios will give managers and investors the ability to undertake a broader and more detailed assessment of companies in the sector, without only focusing on profitability ratios.

## 2. Data and Methodology

Our methodology comprises a number of steps. First, we collect data from financial reports for companies in the semiconductor and related device manufacturing sector. Financial reports are downloaded from the Emerging Markets Information Service (EMIS) database, a Euromoney Institutional Investor Company ([www.emis.com](http://www.emis.com)). The initial sample consists of 2345 companies operating in China (1742), Chinese Taipei (272), South Korea (114), Thailand (48), Singapore (37), India (33), Vietnam (24), Hong Kong (21), Malaysia (20), Russia (10), Turkey (6), Ukraine (4), Ecuador (3) and two companies each from the Czech Republic, Indonesia, Iran, Philippines as well as one each from Bulgaria, El Salvador and Romania. To determine whether investing in companies that manufacture RES components pays off, we divide companies in the sector into two groups. The first group comprises

enterprises that manufacture solar modules, solar cells, solar silicon rods, solar wafers, solar power, solar photovoltaic products and related equipment. We define companies within this group as green companies. The second group comprises companies that are not associated with RES companies. We defined these companies as red companies. The number of companies in our samples is unbalanced that means the number of green companies is much lower than the number of red companies. There are 528 companies in the green group while 1817 companies in the red group. All data are for 2017.

The second step is to construct financial ratios for the companies in our sample using financial reports. Consequently, we consider 62 indicators. Most of these have also been considered in Reference [40]. The ratios considered characterize different aspects of financial performance, namely liquidity, profitability, efficiency, solvency and other aspects (see Table 1). Such variables have been widely used in the analysis of the financial status of companies for the purposes of bankruptcy prediction.

The final step is to apply a decision tree (DT) analysis. We take a similar approach as in bankruptcy prediction models and the principle remains the same; we identify ratios that discriminate between green and red companies. We however do not apply commonly used techniques in bankruptcy prediction. Therefore, we do not use multiple discriminant analysis (MDA) as this technique requires a multivariate normal distribution and equal dispersion matrices. Also, we do not use logit analysis as this method is highly sensitive to multicollinearity and relies upon the assumption of homogenous variation in the data and is highly sensitive to outliers, missing values and extreme non-normality [37,41,42]. Instead, we apply DTs to differentiate between companies associated with the RES manufacturing sector and companies belonging to the general semiconductor and related device manufacturing sector. DT methods have numerous advantages. They require no distributional assumptions or transformed variables. They can handle missing values and qualitative data, are readily understood and can incorporate misclassification costs. Splitting rules are univariate, permitting an easy identification of significant variables. DTs however require the probabilities of successful and failed businesses as inputs. DT algorithms generate a set of tree based classification rules and assign observations to either a successful or failing group, within the context of bankruptcy prediction models. The process begins with a root node, followed by non-leaf nodes which reflect splitting rules—financial ratios. These are then connected to leaf nodes, which represent success or failure. The process of constructing a DT begins with a search for an independent variable that divides observations in a sample in such a way that the difference in the dependent variable is greatest between subgroups. In the next stage, each subgroup is subdivided further by again searching for an independent variable that divides the subgroup so that the difference in the dependent variable is greatest between the subdivided groups. DT algorithms determine the best splitting rule at each non-leaf node. The process continues until splitting no longer produces statistically significant differences in subgroups or subgroups are too small for further division. To simplify the process, some nodes may be removed (pruned), while maintaining a small error rate. DTs differ from logit analysis and MDA in that they identify the relative significance of variables, unlike the latter two approaches which only identify significant variables. DTs nevertheless have predictive power by setting out a sequence of nodes that leads to a classification [36,38,42].

**Table 1.** List of ratios considered.

No	Definition	Group	No	Definition	Group
1	net profit/total assets	P	32	(current liabilities × 365)/cost of products sold	T
2	total liabilities/total assets	S	33	operating expenses/short-term liabilities	S
3	working capital/total assets	L	34	operating expenses/total liabilities	S
4	current assets/short-term liabilities	L	35	profit on sales/total assets	P

Table 1. Cont.

No	Definition	Group	No	Definition	Group
5	$[(\text{cash} + \text{short-term securities} + \text{receivables} - \text{short-term liabilities}) / (\text{operating expenses} - \text{depreciation})] \times 365$	L	36	$(\text{current assets} - \text{inventories}) / \text{long-term liabilities}$	L
6	retained earnings/total assets	P	37	constant capital/total assets	O
7	EBIT/total assets	P	38	profit on sales/sales	P
8	book value of equity/total liabilities	S	39	$(\text{current assets} - \text{inventory} - \text{receivables}) / \text{short-term liabilities}$	L
9	sales/total assets	E	40	$\text{total liabilities} / ((\text{profit on operating activities} + \text{depreciation}) \times (12/365))$	S
10	equity/total assets	S	41	profit on operating activities/sales	P
11	$(\text{gross profit} + \text{extraordinary items} + \text{financial expenses}) / \text{total assets}$	P	42	rotation receivables + inventory turnover in days	L
12	gross profit/short-term liabilities	S	43	$(\text{receivables} \times 365) / \text{sales}$	E
13	$(\text{gross profit} + \text{depreciation}) / \text{sales}$	P	44	net profit/inventory	E
14	EBIT/total costs	S	45	$(\text{current assets} - \text{inventory}) / \text{short-term liabilities}$	L
15	$(\text{total liabilities} \times 365) / (\text{gross profit} + \text{depreciation})$	S	46	$(\text{inventory} \times 365) / \text{cost of goods sold}$	E
16	$(\text{gross profit} + \text{depreciation}) / \text{total liabilities}$	S	47	EBITDA/total assets	P
17	total assets/total liabilities	S	48	EBITDA/sales	P
18	gross profit/total assets	P	49	current assets/total liabilities	S
19	EBIT/sales	P	50	short-term liabilities/total assets	S
20	$(\text{inventory} \times 365) / \text{sales}$	E	51	short-term liabilities/operating expenses	S
21	$\text{sales}(n) / \text{sales}(n-1)$	O	52	equity/fixed assets	O
22	profit on operating activities/total assets	P	53	constant capital/fixed assets	O
23	net profit/sales	P	54	working capital	L
24	gross profit (in 3 years)/total assets	P	55	$(\text{sales} - \text{cost of products sold}) / \text{sales}$	P
25	$(\text{equity} - \text{share capital}) / \text{total assets}$	S	56	net profit/equity	P
26	$(\text{net profit} + \text{depreciation}) / \text{total liabilities}$	S	57	long-term liabilities/equity	S
27	profit on operating activities/financial expenses	E	58	sales/inventory	E
28	working capital/fixed assets	L	59	sales/receivables	E
29	logarithm of total assets	O	60	$(\text{short-term liabilities} \times 365) / \text{sales}$	E
30	$(\text{total liabilities} - \text{cash}) / \text{sales}$	O	61	sales/short-term liabilities	E
31	EBIT/equity	P	62	sales/fixed assets	E

P—represents profitability ratios, L—represents liquidity ratios, S—represents solvency ratios, E—represents efficiency ratios and O—represents other ratios.

To obtain robustness of results, the database is divided into six samples that consist of different numbers of companies and ratios, detailed information can be found in Tables S1–S6 (Supplementary



Materials). From the onset, we consider all 62 indicators and then proceed to reduce this number to 38 and finally to 8 variables. The criterion for removing indicators from our database is a lack of sufficient data for companies comprising a database. The research samples used in the construction of DTs consider a number of companies. Samples with a low number of variables comprise a larger number of companies. The more ratios in a sample, the lower the number of companies it has. The number of enterprises is also lower for samples in which the number of green and red companies is balanced. By reducing the number of indicators, we increase the number of enterprises in the sample. The list of all samples considered in the study together with number of ratios and companies is reported in Table 2.

We view DTs as being well suited for the task at hand. DTs are appropriate for large datasets with a relatively short history. Furthermore, our aim is to build a tree with a minimum number of nodes. Consequently, DT rules are simpler and easier to interpret. The general algorithm comprises the following steps:

1. With a set of  $K$ -records, determine that they belong to the same class. If so, end the algorithm.
2. Otherwise, consider all possible classifications of the overall set  $K$  into subsets  $K_1, K_2, \dots, K_n$  so that they are as homogeneous as possible.
3. Assess each of the classifications according to adopted criteria and select the best one.
4. Divide the set  $K$  according to the adopted criteria.
5. Perform steps 1–4 recursively for each subset.

The subject of division is an  $N$ -element set of objects. Here, these are companies that are characterized by  $M + 1$  ratios (e.g., retained earnings to total assets ratio, (gross profit + extraordinary items + financial expenses) to total assets ratio) which indicate their financial standing. Therefore, the vector  $[y, x]$  together with the respective financial ratios can be defined as follows:

$$[x_n, y_n]_{N \times M+1} = \begin{bmatrix} x_{11} & \dots & x_{1M}y_1 \\ \dots & \dots & \dots \\ x_{N1} & \dots & x_{NM}y_N \end{bmatrix}, \quad (1)$$

where:

$x_1, x_2, \dots, x_M$  are financial variables.

$y$  is the dependent variable, defined a C for red companies and Z for green companies.

Once indicators are defined (data from the Equation (1)), a relationship between  $y$  and the variables  $x_M$  can be established on the basis of the predictors that determine the value of  $y$  as follows:

$$y = f(x, \alpha) + \varepsilon. \quad (2)$$

For this purpose, a recursive split methodology is used to obtain an approximation of the following specification:

$$y = a_0 + \sum_{k=1}^2 a_k I(x \in R_k), \quad (3)$$

where:

$R_k$  for  $k = 1, 2$ —disjoint groups in multidimensional space (red, green),

$a_k$ —model parameters, determined upon the basis of:

$$a_k = \operatorname{argmax}_i \{p(l|k)\}, \quad (4)$$

where:

$p(l|k)$ —probability that an element of the  $R_k$  group belongs to class  $l$ .

**Table 2.** Overview of samples used and the number of financial ratios for each sample.

Sample Name	Number of Ratios	Number of Companies	Number of Red Companies	Number of Green Companies
RG_8V (X1, X7, X9, X19, X21, X23, X24 i X29)	8	2120	1641	479
RG_38V (see Table 3)	38	486	402	84
RG_62V (see Table 1)	62	305	268	37
RG_BD_8V	8	963	484	479
RG_BD_38V	38	188	104	84
RG_BD_62V	62	77	41	37

**Table 3.** Statistical characteristics for all quantitative variables across samples.

Ratio	Min	Max	Average	Standard Deviation	Coefficient of Variation
X2	0.01	3.43	0.42	0.31	0.73
X3	-3.27	0.86	0.24	0.36	1.47
X4	0.046	49.63	3.02	3.83	1.27
X8	-0.71	99.24	3.24	6.22	1.92
X10	-2.43	0.99	0.58	0.31	0.53
X12	-6.00	5.35	0.12	0.78	6.34
X17	0.29	100.24	4.24	6.22	1.47
X18	-4.04	3.09	0.11	0.55	4.91
X20	-7417.70	965.77	54.21	346.95	6.40
X22	-0.53	0.33	0.025	0.10	3.82
X25	-6.62	0.94	0.25	0.57	2.31
X28	-3.88	81.23	1.79	5.05	2.83
X30	-626.60	242.70	0.40	32.44	80.22
X31	-3.26	1.32	0.027	0.33	12.08
X33	-0.003	28.10	3.02	2.73	0.90
X34	-0.003	28.10	2.42	2.48	1.03
X37	-2.43	0.99	0.67	0.28	0.42
X39	0.003	42.74	1.77	3.30	1.86
X41	-7.29	16.07	0.008	0.89	119.71
X42	-54,297.55	1285.76	82.19	2476.86	30.14
X43	-46,879.85	1273.25	27.98	2136.82	76.38
X44	-213.90	40.10	-0.511	10.88	-21.26
X45	0.027	48.86	2.29	3.65	1.59
X46	-11.70	968.97	72.10	73.91	1.03
X47	-0.56	0.37	0.063	0.096	1.53
X48	-5.75	3.79	0.074	0.49	6.65
X49	0.046	26.74	2.39	2.93	1.23
X50	0.009	3.43	0.33	0.28	0.85
X51	-358.67	91.38	0.17	16.88	101.14
X52	-3.22	78.63	2.53	4.92	1.94
X53	-2.88	82.23	2.79	5.05	1.81
X54	-1,884,438.64	16,728,068.12	155,327.35	1,045,613.84	6.73
X56	-3.63	2.13	-0.001	0.35	-304.85
X57	-0.86	3.84	0.23	0.44	1.96
X59	-0.008	275.80	6.25	13.72	2.20
X60	-122,108.85	96,597.68	418.96	8107.69	19.35
X61	-0.003	28.73	3.16	2.81	0.89
X62	-0.009	322.38	4.92	17.93	3.65

RG\_38V, 38 variables, unbalanced number of companies,  $N = 486$ .

The multidimensional space of independent variables ( $X^m$ ) is divided into groups. The model is created by submitting models built in each of  $K$  disjoint groups. For quantitative variables, as used in this study, the model can be stated as follows:

$$I(x \in R_k) = \prod_{m=1}^M I\left(v_{km}^{(d)} \leq x_m \leq v_{km}^{(g)}\right), \quad (5)$$

where:

$v_{km}^{(d)}$ —upper limit of the segment in the  $m$ -th dimension of space,

$v_{km}^{(g)}$ —lower limit of the segment in the  $m$ -th dimension of space,

$I$ —indicator function:

$$I(q) = \begin{cases} 1 & \text{when } q \text{ is true} \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

The DT is a diagrammatic representation of specification (3). The DT algorithm first identifies ratios according to which companies. We term this the training set. It then uses these values to test data set. The model is constructed recursively [43].

We apply three algorithms to construct DTs, namely the Chi-squared Automatic Interaction Detector (CHAID), Classification and Regression Trees (CRT) and Quick, Unbiased, Efficient Statistical Tree (QUEST) algorithms. The CHAID algorithm is an effective algorithm for building DTs developed by Reference [44]. It is mainly used in the segmentation or extension of a DT. The algorithm is appropriate for both quantitative and qualitative variables. It is not a binary method as it can produce more than two categories at any given node. Therefore, this method has the potential to produce a larger DT relative to binary methods. At each step, the CHAID algorithm selects an independent (predictor) variable that has the strongest interaction with the dependent variable. Categories of each predictor are merged if they do not significantly differ in relation to the dependent variable.

The CRT algorithm was developed by Breiman, Firedman, Olshen and Stone in 1984 [45]. Unlike the CHAID algorithm, it is a binary decision algorithm. It is robust to outliers, making it different from other classical methods. It functions in a recurrent manner which means that data is divided into two subsets so that records in each subset are more homogeneous than in the previous sub-set. Both subsets are then again divided until the criterion of homogeneity and other retention criteria are met. The ultimate objective is to maximize homogeneity within sample sub-groups. We apply the Gini index to determine the optimal sub-set division:

$$GI_a(D) = \sum_{c \in C} \frac{|D^c|}{|D|} \left(1 - \frac{|D^c|}{|D|}\right). \quad (7)$$

The QUEST algorithm is a relatively new DT algorithm for binary classifications. It is most often used to classify and explore data [46] and is similar to the CRT algorithm. The difference lies in that the QUEST algorithm is time efficient and unburdened. It does not lose its predictive quality while being efficient and decreases complexity and thereby minimizes DT size [47].

To confirm the accuracy of the resultant classifications, we examined accuracy with various settings for training parameters and apply the 25-fold cross-validation methodology. The calculated risk of cross verification in the output is the risk averaging for 25 test samples. Subsampling DTs are not shown for  $n$ -fold cross verification. Only the DT constructed on the full sample is reported. Similarly, only the full sample classification table is reported. Finally, we also investigate whether based upon the results obtained and the analysis of profitability ratios, green companies exhibit superior performance relative to the remainder of the sector. We point out the most critical indicators occurring in the tree and we test the statistical significance of differences in the selected indicators between groups of companies using the Student's  $t$ -test.

### 3. Results

This section is organized as follows. First, data analysis in defined samples is presented. Then, DT results are shown for outcomes of the 25-fold cross-validation. Next, the analysis of selected of critical indicators is given. Finally, the analysis of profitability ratios is introduced. The outcome of research is presented in Tables 3–26 and Figures 3–5.

#### 3.1. Data Analysis in Defined Samples

We investigate whether the data can be further analyzed and whether the variables are correctly differentiated in the defined samples. Several methods are applied for this purpose. Firstly, basic statistics are calculated for each sample for all variables. These are the minimum value, the maximum value, arithmetic means, standard deviation and variation coefficients. The obtained values exhibit proper characteristics and correctness of data for all analyzed samples in that scope (example calculated values for the database RG\_38V for all companies and for green and red ones are presented in Tables 3–5). For each variable, the variation coefficient exceeds the recommended critical value (higher than 0.1 or 0.15). We can therefore proceed with our analysis.

**Table 4.** Statistical characteristics for all quantitative variables across samples—green companies.

Ratio	Min	Max	Average	Standard Deviation	Coefficient of Variation
X2	0.09	3.43	0.61	0.50	0.82
X3	−3.27	0.82	0.03	0.56	18.22
X4	0.05	7.57	1.66	1.45	0.88
X8	−0.71	10.63	1.39	1.77	1.27
X10	−2.43	0.91	0.39	0.50	1.29
X12	−3.34	2.33	−0.02	0.58	−31.56
X17	0.29	11.63	2.39	1.77	0.74
X18	−2.40	1.09	0.00	0.37	79.84
X20	−7417.70	476.18	−12.47	822.06	−65.92
X22	−0.53	0.24	0.00	0.10	−104.93
X25	−2.62	0.78	0.06	0.61	10.07
X28	−3.88	19.93	0.82	3.06	3.72
X30	−626.60	40.70	−5.32	68.84	−12.94
X31	−3.25	1.32	−0.03	0.52	−16.98
X33	0.05	8.29	1.70	1.50	0.88
X34	0.02	6.57	1.28	1.21	0.94
X37	−2.43	0.94	0.52	0.47	0.91
X39	0.01	4.70	0.68	0.84	1.24
X41	−7.29	16.07	0.03	2.00	57.30
X42	−54,297.55	1285.76	−368.80	5959.94	−16.16
X43	−46,879.85	1273.25	−356.33	5142.79	−14.43
X44	−213.90	18.12	−3.23	23.79	−7.35
X45	0.03	5.77	0.99	1.05	1.06
X46	1.23	530.39	82.55	99.72	1.21
X47	−0.52	0.31	0.03	0.10	3.30
X48	−5.75	3.79	0.00	0.94	−193.69
X49	0.05	6.26	1.24	1.11	0.89
X50	0.06	3.43	0.48	0.47	1.00
X51	0.12	22.20	1.54	3.07	1.99
X52	−3.22	20.58	1.49	3.03	2.04
X53	−2.88	20.93	1.82	3.06	1.68
X54	−1,884,438.64	1,515,112.99	70,859.24	392,833.74	5.54
X56	−3.30	2.14	−0.06	0.59	−9.46
X57	−0.86	2.58	0.36	0.56	1.55
X59	−0.01	275.80	7.31	29.90	4.09
X60	−122,108.85	14,857.05	−698.19	13,558.95	−19.42
X61	0.00	8.75	1.74	1.48	0.85
X62	−0.01	14.92	1.98	2.78	1.41

Green\_38V, 38 variables, green companies, unbalanced number of companies,  $N = 84$ .

**Table 5.** Statistical characteristics for all quantitative variables across samples—red companies.

Ratio	Min	Max	Average	Standard Deviation	Coefficient of Variation
X2	0.01	1.95	0.38	0.23	0.60
X3	−1.51	0.86	0.29	0.28	0.97
X4	0.16	49.63	3.31	4.10	1.24
X8	−0.49	99.24	3.63	6.73	1.85
X10	−0.95	0.99	0.62	0.23	0.37
X12	−6.00	5.35	0.15	0.81	5.32
X17	0.51	100.24	4.63	6.73	1.45
X18	−4.04	3.09	0.13	0.58	4.30
X20	0.00	965.77	68.15	67.74	0.99
X22	−0.30	0.33	0.03	0.09	3.06
X25	−6.62	0.94	0.29	0.55	1.94
X28	−2.67	81.23	1.99	5.35	2.69
X30	−4.66	242.70	1.60	16.85	10.53
X31	−3.26	0.54	0.04	0.26	6.79
X33	0.00	28.10	3.30	2.85	0.86
X34	0.00	28.10	2.66	2.61	0.98
X37	−0.95	0.99	0.70	0.21	0.29
X39	0.00	42.74	2.00	3.56	1.78
X41	−3.96	1.74	0.00	0.37	213.50
X42	8.63	1044.46	176.43	127.43	0.72
X43	0.00	1022.05	108.28	103.57	0.96
X44	−41.72	40.10	0.06	4.91	85.60
X45	0.06	48.86	2.56	3.94	1.54
X46	−11.70	968.97	69.92	67.24	0.96
X47	−0.28	0.37	0.07	0.09	1.35
X48	−3.50	1.62	0.09	0.32	3.58
X49	0.07	26.74	2.63	3.13	1.19
X50	0.01	1.95	0.30	0.21	0.69
X51	−358.67	91.38	−0.12	18.50	−153.02
X52	−1.67	78.63	2.75	5.20	1.89
X53	−1.67	82.23	2.99	5.35	1.79
X54	−1,042,686.22	16,728,068.12	172,977.40	1,135,157.44	6.56
X56	−3.63	0.80	0.01	0.27	23.77
X57	−0.01	3.84	0.20	0.41	2.07
X59	0.00	71.29	6.03	6.50	1.08
X60	12.71	96,597.68	652.39	6413.72	9.83
X61	0.00	28.73	3.45	2.93	0.85
X62	0.00	322.38	5.53	19.62	3.55

Red\_38V, 38 variables, red companies, unbalanced number of companies,  $N = 402$ .

An analysis of the reliability of the variables is also undertaken and we set out to determine whether differentiation in our defined is sufficient for further research. For this purpose, we estimate Cronbach's alpha for the databases and report the results in Table 6. The objective is to eliminate indicators that do not show sufficient differentiation.

The samples with the lowest Cronbach's alpha are the databases constructed using eight indicators. These are the sample with a large disproportion between red and green records (RG\_8V) and the balanced sample (RG\_BD\_8V). However, both values indicate the viability for further analysis. The remaining four samples exhibit sufficiently high variable differentiation. The largest differentiation occurs in the sample comprising a balanced number of records of red and green companies for all 62 indicators (RG\_BD\_62V). A slightly lower value is obtained for the database comprising all 62 indicators (RG\_62V) but without a balanced number of records for both groups of companies. The remaining two groups (RG\_38V, RG\_BD\_38V) show slightly lower but nevertheless sufficient differentiation. Our Cronbach's alpha estimates confirm the viability of further analysis for the groups of variables in all samples.

**Table 6.** Cronbach's alpha across samples.

Database Name	Cronbach's Alpha
RG_8V	0.467
RG_38V	0.796
RG_62V	0.820
RG_BD_8V	0.380
RG_BD_38V	0.797
RG_BD_62V	0.857

### 3.2. DT Results for Whole Companies

The first samples include 38 indicators. It comprises 486 companies, of which 402 are red companies and the remaining 84 are green companies. To limit the size of our DTs in the beginning, the minimum number of observations in the parent node is defined as 100 and the child node as 50. The algorithm identifies four independent variables, namely the short-term liabilities to operating expenses ratio (X51), days sales of inventory II ratio (X46), days sales of inventory I ratio (X20) and the debt to assets ratio (X2), to construct the DT. The DT subsequently comprises nine nodes, including five end nodes. Unfortunately, the classification matrix following the application of 25-fold cross-validation indicates that it is not possible to correctly classify companies in relevant groups. The effect of cross-validation for various numbers of companies in parent and child nodes is presented in Table 5. The algorithms identify financial ratios which unambiguously show that a company manufactures RES component or not. Next, we reduced the number of companies in the parent and child node to 50 and then 25, followed by a reduction to 20 and finally to 10. The reduction in the number of companies in the nodes improves classification ability attributable to the algorithms. The effect of cross-validation for a different number of records at parent and child nodes is reported in Table 7.

**Table 7.** Cross-validation classification matrix for 25-fold cross-validation for sample of RG\_38V and decision tree (DT) algorithm: CHAID.

The Number of Records in the Parent and Child Nodes: 100 and 50			
Observed	predicted		percentage correct
	C	Z	
C	402	0	100.0%
Z	84	0	0.0%
total percentage	100.0%	0.0%	82.7%
The Number of Records in the Parent and Child Nodes: 50 and 25			
Observed	predicted		percentage correct
	C	Z	
C	376	26	93.5%
Z	51	33	39.3%
total percentage	87.9%	12.2%	84.2%
The Number of Records in the Parent and Child Nodes: 20 and 10			
Observed	predicted		percentage correct
	C	Z	
C	366	36	91.0%
Z	44	40	47.6%
total percentage	84.4%	15.6%	83.5%

We note that the short-term liabilities to operating expenses (X51) ratio, is the most important ratio, according to which the first classification in the DT is established. The second one is X46 days sales of inventory II. Cross-validation for the CRT algorithm is at 99.0% for red and 20.2% for green companies respectively and the QUEST algorithm achieves a cross-validation level of 97.5% for red companies and 19.0% for green companies, respectively. According to the QUEST algorithm, the



first classifying ratio is the equity to total asset ratio (X10). The QUEST algorithm underperforms the CHAID algorithm, as reported in Table 8.

**Table 8.** Cross-validation classification matrix for the RG\_38V sample for the three DT algorithms.

Algorithm	Red	Green	First Division Variable
CHAID	91.0%	47.6%	X51
CRT	99.0%	20.2%	X51
QUEST	97.5%	19.0%	X10

Sample with dependant variable as colour (C—red enterprises, Z—green enterprises). Number of records in parent and child nodes: 20 and 10.

A further consideration is whether a reduction in the number of indicators coupled with a simultaneous increase in the number of companies in the sample will positively impact the classification capability of DT algorithms. A reduced number of indicators permits the definition of a sample with a higher number of companies. The sample of only eight ratios comprises 2120, of which 479 are green companies.

The classification option for the sample constructed in this way has been checked using all three DT algorithms. The materiality level for division nodes is set at the level of 0.05 permitting a maximum number of 100 iterations and Pearson's chi-square statistics are applied. DT extension is limited to a maximum of three levels. In application of DTs to the RG\_8V sample, cross-validation shows very poor accuracy correctness. All three DT algorithms are applied to 20 companies in the parent node and 10 in the child node (Table 9).

**Table 9.** Cross-validation classification matrix for the database RG\_8V.

Algorithm	Red	Green	First Division Variable
CHAID	99.1%	4.6%	X21
CRT	97.9%	13.6%	X21
QUEST	100.0%	0.0%	X21

Three algorithms for DTs with, dependent variable; colour (C—red enterprises, Z—green enterprises), number of records in parent and child nodes: 20 and 10.

Next, we construct a sample for all indicators which is accompanied by a decrease in the number of companies. This sample comprises 62 variables and 305 companies, including 37 records green companies which means 12.13% of all companies in the sample. The number of records of parent and child nodes has been reduced to proportions of 20 and 10. We leave DT size unchanged at the third level as no material changes are observed in the other levels. Following the reduction in the number of records in parent and child nodes, the CHAID algorithm produces a correct classification rate of 96.6% for red companies and 29.7% for green companies. Of all DT algorithms, the CHAID algorithm produces the best results whereas the CRT algorithm underperforms classifying 96.6% of red companies and 27.0% of green companies correctly. The QUEST algorithm does not show any classification capability for green companies (see Table 10).

**Table 10.** Cross-validation classification matrix for database RG\_62V database.

Algorithm	Red	Green	First Division Variable
CHAID	96.6%	29.7%	X5
CRT	96.6%	27.0%	X50
QUEST	100.0%	0.0%	No variable

Three algorithms, CHAID, CRT and QUEST algorithms applied for DT construction. The dependant variables are green and red companies. The number of records in parent and child nodes is set to 20 and 10 respectively.

Based upon these results, the question that can be posed is as to whether a reduction in the disproportion between the number of records for green and red companies respectively in the databases will have a material impact on the classification capacity of the algorithm. Consequently, we balance the number of companies in all databases under consideration. The balanced database with the main eight indicators (RG\_BD\_8V) comprises 963 companies of which 479 constitute green enterprises. The balanced database with 38 indicators (RG\_BD\_38V) comprises 188 companies, of which 84 are green enterprises. The database with the highest number of indicators (RG\_BD\_62V)\_ comprises only 77 companies, of which 37 are green companies. The results of this classification are presented in Table 13.

Similar to the databases with an unbalanced and balanced number of green and red companies with 38 indicators (RG\_38V & RG\_BD\_38V), the classifying ratio according to which the first division the DT is undertaken is X51. Balancing records in this database brings the expected effect in the form of a material improvement in the DT classification of companies. The number of correctly classified companies now exceeds 50%. The CHAID algorithm produces the most favourable results, with the number of records in the parent and child nodes equalling 20 and 10 and an accurate classification rate of 84.6% for red companies and 67.9% for green enterprises (Table 11).

**Table 11.** Cross-validation classification matrix for the RG\_BD\_38V sample using the CHAID algorithm.

<b>The Number of Records in the Parent and Child Nodes: 100 and 50</b>			
Observed	predicted		percentage correct
	C	Z	
C	81	23	77.9%
Z	32	52	61.9%
total percentage	60.1%	39.9%	70.7%
<b>The Number of Records in the Parent and Child Nodes: 50 and 25</b>			
Observed	predicted		percentage correct
	C	Z	
C	97	7	93.3%
Z	46	38	45.2%
total percentage	76.1%	23.9%	71.8%
<b>The Number of Records in the Parent and Child Nodes: 20 and 10</b>			
Observed	predicted		percentage correct
	C	Z	
C	88	16	84.6%
Z	27	57	67.9%
total percentage	61.2%	38.8%	77.1%

The remaining three algorithms also show material improvements in classification ability. The CRT algorithm, in contrast to CHAID, yields a more accurate classification rate for green companies with 71.4% of companies classified correctly and 81.7% of red companies classified correctly, somewhat fewer than the CHAID algorithm. The QUEST algorithm underperforms, classifying 74.0% of red companies and 58.3% of green companies correctly. In all three cases, the variable used in the first classification is the short-term liabilities to operating expenses ratio (X51) Results are summarized in Table 12.

**Table 12.** Cross-validation classification matrix for the RG\_BD\_38V sample.

Algorithm	Red	Green	First Division Variable
CHAID	84.6%	67.9%	X51
CRT	81.7%	71.4%	X51
QUEST	74.0%	58.3%	X51

Three algorithms of tree building, dependant variable: colour (C—red enterprises, Z—green enterprises), number of records in parent and child nodes: 20 and 10.

In the balanced database comprising eight indicators, the reduction in the maximum number of companies in parent and child DT nodes from approximately 100 to 50 and then from 50 to 25, improves classification accuracy, as evident in Table 13. Although the percentage of correctly classified green companies increases from 57.6% to 63.3%, at the same time, the percentage of correctly classified red companies decreases from 66.1% to 61.6%. However, the resultant DT is now larger, with three levels of depth and as many as 12 nodes. This contrasts with 10 nodes in the unbalanced database.

**Table 13.** Cross-validation classification matrix for the RG\_BD\_8V sample, tree-building algorithm: CHAID.

The Number of Records in the Parent and Child Nodes: 100 and 50			
Observed	predicted		percentage correct
	C	Z	
C	320	164	66.1%
Z	203	276	57.6%
total percentage	54.3%	45.7%	61.9%
The Number of Records in the Parent and Child Nodes: 50 and 25			
Observed	predicted		percentage correct
	C	Z	
C	298	186	61.6%
Z	176	303	63.3%
total percentage	49.2%	50.8%	62.4%
The Number of Records in the Parent and Child Nodes: 20 and 10			
Observed	predicted		percentage correct
	C	Z	
C	298	186	61.6%
Z	176	303	63.3%
total percentage	49.2%	50.8%	62.4%

The three years gross profit to asset ratio (X24) is the first classifying ratio in the DT. For the unbalanced database with eight indicators (RG\_8V), revenue growth ratio (X21) is the classification ratio the root. Other DT algorithms also show higher correct classification rates. The best classification accuracy for green companies is attained by the CRT algorithm, which correctly classifies 61.2% of red companies and 70.1% of green companies. The classification matrix for all DT algorithms for the balanced database with eight indicators is presented in Table 14.

**Table 14.** Cross-validation classification matrix for the database RG\_BD\_8V.

Algorithm	Red	Green	First Division Variable
CHAID	61.6%	63.3%	X24
CRT	61.2%	70.1%	X24
QUEST	60.1%	59.9%	X29

Four algorithms of tree building, dependant variable: colour (C—red enterprises, Z—green enterprises), number of records in parent and child nodes: 20 and 10.

The balanced database comprising all 62 indicators (RG\_BD\_62V) comprises the lowest number of records. It is composed of data for 37 green enterprises and 41 red enterprises, comprising a total of 77 records. The number of records in the parent node is set to 50 and then 20 and in the child node to 25 and 10 respectively. The CHAID algorithm produces the best results (Table 15). It correctly classifies 70.3% of green enterprises and 85.0% of red enterprises. The QUEST algorithm underperforms all other algorithms, failing to correctly classify any green enterprises (Table 16).

**Table 15.** Cross-validation classification matrix for the RG\_BD\_62V database, tree-building algorithm: CHAID.

The Number of Records in the Parent and Child Nodes: 50 and 25			
Observed	predicted		percentage correct
	C	Z	
C	27	13	67.5%
Z	12	25	67.6%
total percentage	50.6%	49.4%	67.5%
The Number of Records in the Parent and Child Nodes: 20 and 10			
Observed	predicted		percentage correct
	C	Z	
C	34	6	85.0%
Z	11	26	70.3%
total percentage	58.4%	41.6%	77.9%

**Table 16.** Cross-validation classification matrix for the RG\_BD\_62V database.

Algorithm	Red	Green	First Division Variable
CHAID	85.0%	70.3%	X3
CRT	75.0%	67.6%	X6
QUEST	100.0%	0.0%	Variables not considered

Three algorithms of tree building, dependant variable: colour (C—red enterprises, Z—green enterprises), number of records in parent and child nodes: 20 and 10.

Using a balanced number of records for red enterprises and green enterprises in all three databases translates into increased classification accuracy for green enterprises. Concurrently, the classification accuracy decreases somewhat for red enterprises (see Table 17).

**Table 17.** Percentage of correctly classified enterprises based upon cross-validation for all analysed samples.

The Number of Records in the Parent and Child Nodes: 100 and 50						
	RG_BD_8V	RG_8V	RG_BD_38V	RG_38V	RG_BD_62V	RG_62V
C	66.1%	100.0%	77.9%	100.0%	—	100.0%
Z	57.6%	0.0%	61.9%	0.0%	—	0.0%
The Number of Records in the Parent and Child Nodes: 50 and 25						
	RG_BD_8V	RG_8V	RG_BD_38V	RG_38V	RG_BD_62V	RG_62V
C	61.6%	99.1%	93.3%	93.5%	67.5%	100.0%
Z	63.3%	4.6%	45.2%	39.3%	67.6%	0.0%
The Number of Records in the Parent and Child Nodes: 20 and 10						
	RG_BD_V	RG_8V	RG_BD_38V	RG_38V	RG_BD_62V	RG_62V
C	61.6%	99.1%	84.6%	91.0%	85.0%	96.6%
Z	63.3%	4.6%	67.9%	47.6%	70.4%	29.7%

Tree-building algorithm: CHAID, dependant variable: colour (C, Z).

Using the balanced database (RG\_BD\_8V) for eight variables, the percentage of correctly classified green enterprises for the database increases from 4.6% to 63.3% and for 38 variable database, the percentage of correctly classified green enterprises increases from 47.6% to 67.9% when the databased is balanced. For the database of 62 variables, it increases from 29.7% to 70.4%.

An analogous our study of classification using decision trees was also carried out for grouped indicators. As previous studies show that only balanced databases give positive effects for the classification of enterprises into green and red (see Table 17), only these databases were used for the study, divided by groups of indicators (see Table 18).

**Table 18.** Overview of samples used and the number of financial ratios for each sample.

Sample Name	Number of Variable	Groups and Indicators
RG_BD_8V	8	Group P: X1, X7, X19, X23, X24
RG_BD_38V	38	Group D: X2, X8, X10, X12, X17, X25, X33, X34, X49, X50, X51, X57; Group L: X3, X4, X28, X39, X45, X54; Group T: X20, X42, X43, X44, X46, X59, X61, X62
RG_BD_62V	62	All indicator groups, all indicators

Databases have been built to analyse the indicators indicated in the groups. The indicators included in the group and the indicators used in research to analyse the RG\_BD\_62V and RG\_BD\_38V databases are presented in Table 19.

**Table 19.** Indicators included in the group and those used in the analysis of the RG\_BD\_62V and RG\_BD\_38V databases.

Group	RG_BD_62V
P:	X1, X6, X7, X11, X13, X18, X19, X22, X23, X24, X31, X35, X38, X41, X47, X48, X55, X56,
S:	X2, X8, X10, X12, X14, X15, X16, X17, X25, X26, X33, X34, X40, X49, X50, X51, X57,
L:	X3, X4, X5, X28, X36, X39, X45, X54,
E:	X9, X20, X32, X42, X43, X44, X46, X58, X59, X60, X61, X62,
O:	X21, X27, X29, X30, X37, X52, X53,
Group	RG_BD_38V
S:	X2, X8, X10, X12, X17, X25, X33, X34, X40, X49, X50, X51, X57,
L:	X3, X4, X28, X39, X45, X54,
E:	X20, X42, X43, X44, X46, X59, X61, X62,

Two balanced databases, namely the RG\_BD\_38V and RG\_BD\_62V databases, were used for calculations. The most interesting results are presented. Our studies have shown that the CHAID method exhibits the best classification performance. This method was used in this analysis. For group P indicators—the profitability indicators—cross-validation showed a correct classification rate of 85 percent for red enterprises and 70.3 percent for green enterprises (Table 20). The use of cross-validation for a database with fewer indicators (RG\_BD\_38V) resulted in 88.5 percent of correctly classification red enterprises and 51.2percent of green enterprises. The result is slightly worse for the database with 62 indicators. However, for the group of E indicators, namely efficiency indicators, the result presents an improvement when using a database with 38 indicators (70.2 percent—red, 73.8 percent—green) compared to a database with 62 indicators (95.0 percent—red but only 35.1 percent—green).

**Table 20.** Cross-validation classification matrix for the RG\_BD\_38V sample and the RG\_BD\_62V sample for the DT algorithm for indicator groups.

Groups of Indicators	Red	Green	First Division Variable
<b>RG_BD_62V</b>			
Group P	85.0%	70.3%	X55
Group S	85.0%	45.9%	X40
Group L	95.0%	35.1%	X3
Group E	100.0%	0.0%	No variable is included
Group O	95.0%	35.1%	X52
<b>RG_BD_38V</b>			
Group S	88.5%	51.2%	X33
Group L	75.0%	58.3%	X4
Group E	70.2%	73.8%	X46

Sample with dependant variable as colour (C—red enterprises, Z—green enterprises).

Following the removal of a single outlier, the model fit improves significantly. Data on green enterprises from the RG\_38V database were used to build a linear regression model. The model was built using the stepwise method. Initially, all 38 variables constituting the examined database were introduced into the model (see Table 3). The least squares method was used to fit a regression through a set of observations. The level of significance for retaining a variable was set at 0.05 and at 0.1 for deletion. The dependent variable is X56 (net profit/equity). Following this variable selection procedure, a model comprising 10 predictors was obtained. This model has the highest value of the coefficient of determination R-squared ( $R^2 = 0.990$ ,  $R^2 - corrected = 0.988$ ). The representative equation is as follows:

$$y = -0.126 + 1.006 \times X31 - 0.356 \times X48 + 0.262 \times X41 - 0.00002 \times X60 + 1.49 \times X47 - 0.088 \times X57 - 1.310 \times X22 - 0.025 \times X34 + 0.167 \times X37 + 0.022 \times X51. \quad (8)$$

All coefficients are statistically significant (Table 21) and the model exhibits the desired statistical significance (Table 22).

**Table 21.** Statistical significance of indicators.

Model	Variable	Significance
(Constant)		0.002
X31	EBIT/equity	0.000
X48	EBITDA/sales	0.000
X41	profit on operating activities/sales	0.000
X60	sales	0.000
X47	EBITDA/total assets	0.000
X57	long-term liabilities/equity	0.000
X22	profit on operating activities/total assets	0.018
X34	operating expenses/total liabilities	0.001
X37	constant capital/total assets	0.000
X51	short-term liabilities/operating expenses	0.015

Endogeneity regarding correlated-omitted-variables bias for the built regression model has been tested. For this purpose, the impact threshold for a confounding variable (ITCV) test was used [48,49]. An impact statistic were calculated (ITCV) indicating the minimum impact of a confounding variable that would be needed to render the coefficient statistically insignificant. According to the received value of ITCV,  $\alpha = 0.05$ , an omitted variable would have to be correlated at 0.983 with the outcome and at 0.983 with the predictor of interest (conditioning on observed covariates) to invalidate an



inference based on a threshold of 0.23 for statistical significance. Correspondingly the impact of an omitted variable must be  $0.966$  ( $0.983 \times 0.983$ ) to invalidate an inference.

**Table 22.** Statistical significance of the regression model.

Model	Anova <sup>a</sup>				
	Sum of Squares	df	Mean Square	F	Significance
Regression	28.123	10	2.812	694.839	0.000
Rest	0.295	73	0.004		
Altogether	28.419	83			

<sup>a</sup> Dependent variable: X56. Predictors: (Constant), X31, X48, X41, X60, X47, X57, X22, X34, X37, X51.

### 3.3. Case Study for Data only for Chinese Companies

Basic research was carried out for 2345 enterprises from various countries, in particular for Asian countries (over 93%). All data was for 2017. First, estimations for all companies were made and they became the basis for calculations for which only Chinese companies were selected from the database. The data constitutes almost 75% of the entire database. The database configured in this way was used to build two databases for enterprises in China, one with 8 indicators and the other with 38 indicators. A database of financial results of Chinese enterprises was used for the calculations. Initially, the database comprised 1742 observations in total (almost 75% of the previous database with data from several different countries), of which 1267 were red enterprises and 475—green enterprises. Because previous studies have shown that the decision tree algorithm works best on databases with an equal number of green and red enterprises, the database has been balanced. Finally, due to lack of data, two databases for Chinese enterprises with different numbers of indicators (8 and 38) and balanced but also different numbers of observations were built (Table 23).

**Table 23.** Overview of samples used and the number of financial ratios for each sample.

Sample Name	Number of Ratios	Number of Companies	Number of Red Companies	Number of Green Companies
China_RG_BD_8V (X1, X7, X9, X19, X21, X23, X24 i X29)	8	843	422	421
China_RG_BD_38V (see Table 3)	38	115	83	32

Calculations have been made for these two balanced databases. The result of the classification is similar to the results for the database in which enterprises were mixed from several countries (Table 24). This is due to the fact that data for Chinese companies constitutes the majority in the analysed database (almost 75%). The first division variables were X21 and X8 respectively.

**Table 24.** Percentage of correctly classified enterprises based upon cross-validation for analysed samples.

	China_RG_BD_8V	RG_BD_8V	China_RG_BD_38V	RG_BD_38V
C	69.2%	61.6%	72.7%	84.6%
Z	55.8%	63.3%	71.9%	67.9%

Tree-building algorithm: CHAID, dependant variable: colour (C, Z).

### 3.4. Analysis of Critical Indicators

The use of DTs makes it possible to identify important indicators that distinguish companies within the sector. The list of our classifying ratios is presented in Table 25. Five ratios represent different groups of ratios, relating to liquidity, turnover, one from our initial analysis and profitability.

**Table 25.** Summary of the most important indicators.

Ratio	Definition	Classification Threshold	Group
X51	short-term liabilities/operating expenses	$\leq 0.84389330637$ ; $> 0.84389330637$ $\leq 0.69232544490$ ; $> 0.69232544490$	Efficiency
X21	sales ( $n$ )/sales ( $n - 1$ )	$\leq 0.92510251238$ ; $> 0.92510251238$	Other
X5	$[(\text{cash} + \text{short-term securities} + \text{receivables} - \text{short-term liabilities})/(\text{operating expenses} - \text{depreciation})] \times 365$	$\leq -154.4882205507$ ; $> -154.4882205507$	Liquidity
X24	gross profit (in 3 years)/total assets	$\leq 0.101199752312$ ; $> 0.101199752312$	Profitability
X3	working capital/total assets	$\leq -0.04734495587$ ; $> -0.04734495587$	Liquidity

We identify the short-term liabilities to operating expenses ratio (X51) as the most important ratio for classification. The second most important ratio is the revenue growth ratio (X21). On the other hand, the last one is the size of working capital ratio. This ratio is also used in bankruptcy prediction models for manufacturing companies and its inclusion is motivated by Altman's (1968) model [32]. In Table 26, the differences between the estimated values of financial ratios associated with green and red companies are tested.

**Table 26.** The Student's  $t$ -test results for each sample for most important ratios.

Ratio	RED	GREEN	$t$ -Value	df	$p$
<b>RG_8V</b>					
X21	1.07	0.99	4.26	2017	0.000021 ***
X24	0.11	0.06	4.90	1936	0.000001 ***
<b>RG_BD_8V</b>					
X21	1.06	0.99	2.41	908	0.016344 **
X24	0.11	0.06	4.90	885	0.000001 ***
<b>RG_38V</b>					
X3	0.30	0.13	5.53	475	0.000000 ***
X51	0.43	0.65	-6.04	434	0.000000 ***
<b>RG_BD_38V</b>					
X3	0.27	0.12	4.13	181	0.000055 ***
X51	0.52	0.81	-4.72	174	0.000005 ***
<b>RG_62V</b>					
X3	0.30	0.17	2.93	257	0.003719 ***
X5	47.28	-10.98	2.30	257	0.022226 **
X21	1.07	1.02	1.22	257	0.225125 *
X24	0.06	-0.06	2.54	257	0.011714 **
X51	0.39	0.49	-2.28	257	0.023180 **
<b>RG_BD_62V</b>					
X3	0.27	0.14	2.30	66	0.024833 **
X5	40.51	-37.98	2.19	66	0.032349 **
X21	1.05	1.05	0.11	66	0.912253 *
X24	0.098	-0.05	2.49	66	0.015349 **
X51	0.41	0.54	-2.18	66	0.033077 *

\* indicates statistical significance at 10%, \*\* indicates statistical significance at 5% and \*\*\* indicates statistical significance at 1%.

Aside from the revenue growth ratio (for two databases, RG\_62V and RG\_BD\_62V), these results confirm that the differences in ratios are statistically significant. This confirms that these ratios are important for classifying companies belonging to the semiconductor and solid state device

manufacturing sector. Moreover, green companies that produce components for renewable energy sources (RES) are characterized by lower financial ratios except the payables to operating expenses ratio, which indicate generally higher levels of debt.

### 3.5. Analysis of Profitability Ratios

The analysis of selected indicators shows that only one profitability indicator has been identified as critical. Therefore, key indicators from the investor and manager’s point of analyzed in this subsection are the namely the return on assets (ROA), the return on sales (ROS) and the return on equity (ROE) ratios. It is worth adding that the selected profitability ratios were calculated for China given its representativity in the sample. Three-quarters of the sample comprises companies operate in China. Figures 3–5 also confirm that green companies are described by lower profitability ratios, which are the ROA (X7), ROS (X19) and ROE (X31) ratios.

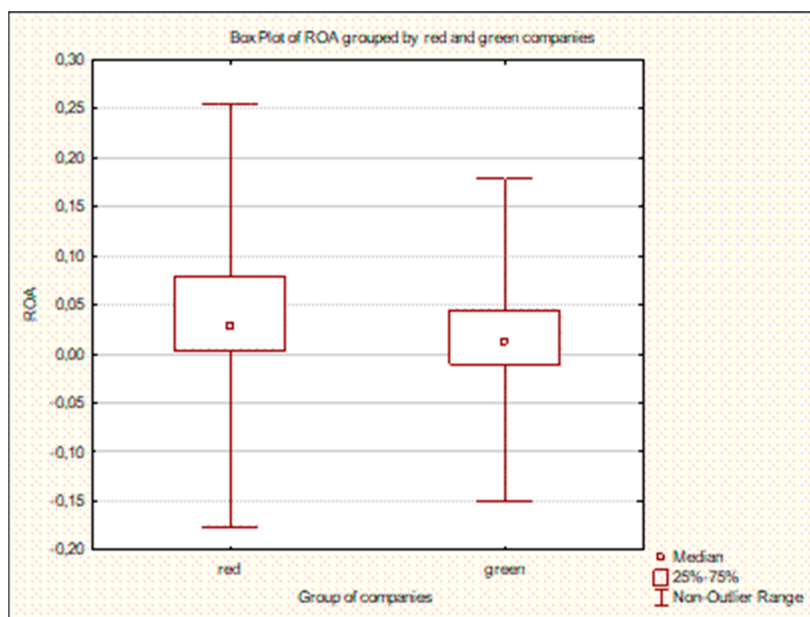


Figure 3. Return on assets (ROA) values for red and green companies.

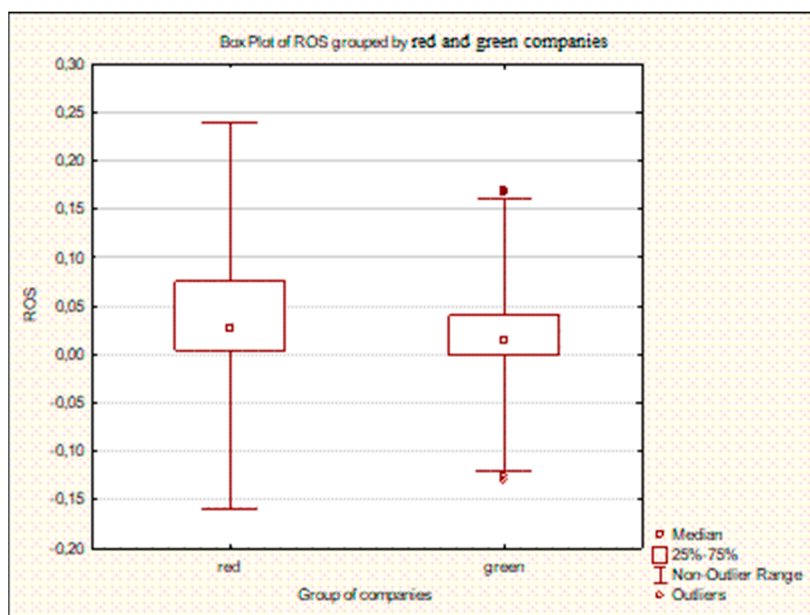


Figure 4. Return on sales (ROS) values for red and green companies.

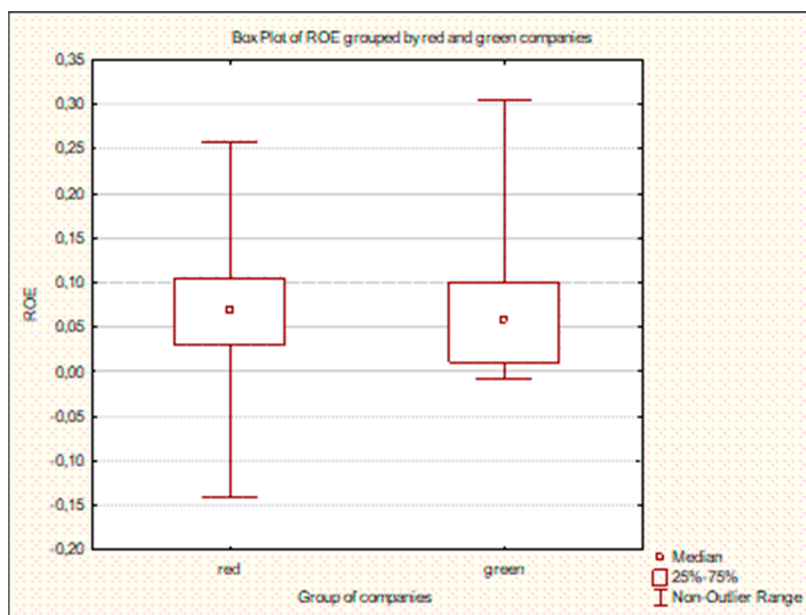


Figure 5. Return on equity (ROE) values for red and green companies.

#### 4. Conclusions

Our study investigates whether investing in the solar technology manufacturers' components is a lucrative business by analysing the semiconductor and related device manufacturing sector. To do so, we apply a novel approach for the sector, namely DTs. We consider 62 financial ratios and our initial database comprises 2345 companies, mostly operating in China. The companies in the sector are classified into two groups, green companies which are those for which production is related to renewable energy and red companies which are those for which production is not related to renewable energy. We define six samples, both balanced and unbalanced samples and comprising large and small samples of companies and 25-fold cross-validation, which does not significantly improve the results. The literature reports on RES companies, with most of the existing literature related to power plants [21,22], electric utilities [24] or energy companies [30]. Our results are similar to other studies of RES companies. On the basis of certain financial ratios, we find that green companies underperform red companies [21,22]. Our contribution lies in addressing the lack of studies that assess the relative financial performance of companies that provide equipment needed for renewable energy production. Moreover, we apply DTs for the purposes of identifying classifying ratios.

The decision tree based analysis identifies the most important indicators for evaluating enterprises, especially for those operating in China. This provides a broader overview of the financial standing of groups of companies to managers and investors. Our results indicate that investing in companies that manufacture RES components may not be a lucrative business. For managers, this is important information that indicates that RES companies are not profitable or as financially sound relative to companies in the general semiconductor and solid state device manufacturing sector. For investors potentially interested in investing in RES associated companies, our findings demonstrate a need for caution. This should be viewed within the context of achieving RES targets in total energy production. Given government targets for renewable energy, the demand for components for renewable energy production will increase. However, it remains to be seen whether this increase in demand will accompany profits.

The topic chosen by us for analysis is difficult and multi-threaded but it is worth making a foundation for further research. Our research provides a solid basis for further exploration of this data. With this foundation in place, we can continue research in the directions indicated by the reviewer. The authors are aware that all factors, both fundamental and economic, may impact the profitability of doing business.

Our analysis, however, focuses on financial ratios and is a good basis for further research. Specifically, we focus on companies within the semiconductor and related device manufacturing sector. The purpose of our work is not to compare which industry is more profitable for the investor but to determine whether selected financial indicators will indicate differences between enterprises related to the production of semi-finished products for renewable energy production and conventional energy.

Given the importance of renewable energy sources for investors, information on the profitability of not only these enterprises themselves but also enterprises in the entire production chain is crucial. Our research is significant and fills the gap in the analysed area regarding the analysis of financial indicators for enterprises from the semiconductor and related device manufacturing sector.

We recognize some of the limitations of our study. First, the research period only encompasses 2017. This may influence results. We are faced with this limitation owing to missing data for the years 2008–2016. Second, three-quarters of analysed enterprises in our sample are from China. Third, we only rely upon financial ratios in our study. Finally, we limit ourselves to applying only DTs in this study. Other approaches, such as neural network (NN) can also be used for classification purposes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1996-1073/13/2/499/s1>. File consists of Tables S1–S6. Table S1. RG\_8V; 8 variables, unbalanced number of companies, N = 2120. Table S2. RG\_BD\_8V; 8 variables, balanced number of companies, N = 963. Table S3. RG\_38V; 38 variables, unbalanced number of companies, N = 486. Table S4. RG\_BD\_38V, 38 variables, balanced number of companies, N = 188. Table S5. RG\_62V, 62 variables, unbalanced number of companies, N = 305. Table S6. RG\_BD\_62V; 62 variables, balanced number of companies, N = 77. ID concerns red enterprises (0) and green enterprises (1); colour concerns red enterprises (C) and green enterprises (Z).

**Author Contributions:** Conceptualization, S.K.T., A.S.-S.; and J.J.S.; methodology, S.K.T., A.S.-S.; software, S.K.T., A.S.-S.; validation, S.K.T., A.S.-S.; formal analysis, S.K.T., A.S.-S.; investigation, S.K.T., A.S.-S.; resources, S.K.T., A.S.-S.; data curation, S.K.T.; writing—original draft preparation, S.K.T., A.S.-S., J.J.S.; writing—reviewing and editing, S.K.T., J.J.S.; visualization, S.K.T.; supervision, S.K.T.; project administration, S.K.T.; funding acquisition, S.K.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Wrocław University of Science and Technology, financed from statutory funds, project No. 0402/0135/18.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. The International Energy Agency. Available online: <https://www.iea.org/statistics> (accessed on 30 October 2019).
2. Independent Evaluation Group. CHINA Renewable Energy Scale-Up Program: Phase One, Report No. 117156. 2017. Available online: <https://ieg.worldbankgroup.org/sites/default/files/Data/ppar-chinarenewableenergy-10302017.pdf> (accessed on 22 December 2019).
3. National Institute of Transforming India. Report of Expert Group on 175 GW RE by 2022. 2015. Available online: <https://niti.gov.in/writereaddata/files/175-GW-Renewable-Energy.pdf> (accessed on 22 December 2019).
4. The International Energy Agency. Available online: <https://www.iea.org/policiesandmeasures/renewableenergy> (accessed on 30 October 2019).
5. The International Energy Agency. Available online: <https://www.iea.org/policiesandmeasures/pams/china> (accessed on 30 October 2019).
6. Sinha, A. Here Are India's INDC Objectives and How Much It Will Cost. Available online: <https://indianexpress.com/article/india/india-news-india/here-are-indias-indc-objectives-and-how-much-it-will-cost/> (accessed on 30 October 2019).
7. Hongzhan, S.; Qiang, Z.; Yibo, W.; Qiang, Y.; Jun, S. China's solar photovoltaic industry development: The status quo, problems and approaches. *Appl. Energy* **2014**, *118*, 221–230.
8. Pablo-Romero, M.P. Solar Energy: Incentives to Promote PV in EU27. *AIMS Energy* **2013**, *1*, 28–47. [[CrossRef](#)]
9. Solangi, K.H.; Islam, M.R.; Saidur, R.; Rahim Fayaz, H. A review on global energy policy. *Renew. Sustain. Energy Rev.* **2011**, *15*, 2149–2163. [[CrossRef](#)]

10. Sachu, B.K. A study on global solar PV energy developments and policies with a special focus on the top ten solar PV power producing countries. *Renew. Sustain. Energy Rev.* **2015**, *43*, 621–634.
11. Trück, S.; Weron, R. Convenience yields and risk premiums in the EU-ETS—Evidence from the Kyoto commitment period. *J. Futures Mark.* **2016**, *36*, 587–611. [[CrossRef](#)]
12. Bohl, M.T.; Kaufmann, P.; Stephan, P.M. From hero to zero: Evidence of performance reversal and speculative bubbles in German renewable energy stocks. *Energy Econ.* **2013**, *37*, 40–51. [[CrossRef](#)]
13. Henriques, I.; Sadorsky, P. Oil prices and the stock prices of alternative energy companies. *Energy Econ.* **2008**, *30*, 998–1010. [[CrossRef](#)]
14. Inchauspe, J.; Ripple, R.D.; Truck, S. The dynamics of returns on renewable energy companies: A state-space approach. *Energy Econ.* **2015**, *48*, 325–335. [[CrossRef](#)]
15. Kumar, S.; Managi, S.; Matsuda, A. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Econ.* **2012**, *34*, 215–226. [[CrossRef](#)]
16. Managi, S.; Okimoto, T. Does the price of oil interact with clean energy prices in the stock market? *Jpn. World Econ.* **2013**, *27*, 1–9. [[CrossRef](#)]
17. Maryniak, P.; Trück, S.; Weron, R. Carbon pricing and electricity markets—The case of the Australian Clean Energy Bill. *Energy Econ.* **2019**, *79*, 45–58. [[CrossRef](#)]
18. Capece, G.; Cricelli, L.; Di Pillo, F.; Levialdi, N. A cluster analysis study based on profitability and financial indicators in the Italian gas retail market. *Energy Policy* **2010**, *38*, 3394–3402. [[CrossRef](#)]
19. Capece, G.; Cricelli, L.; Di Pillo, F.; Levialdi, N. New regulatory policies in Italy: Impact on financial results, on liquidity and profitability of natural gas retail companies. *Util. Policy* **2012**, *23*, 90–98. [[CrossRef](#)]
20. Halkos, G.E.; Tzeremes, N.G. Analyzing the Greek renewable energy sector: A Data Envelopment Analysis approach. *Renew. Sustain. Energy Rev.* **2012**, *16*, 2884–2893. [[CrossRef](#)]
21. Paun, D. Sustainability and financial performance of companies in the energy sector in Romania. *Sustainability* **2017**, *9*, 1722. [[CrossRef](#)]
22. Tomczak, S.K. Comparison of the Financial Standing of Companies Generating Electricity from Renewable Sources and Fossil Fuels: A New Hybrid Approach. *Energies* **2019**, *12*, 3856. [[CrossRef](#)]
23. Clarkson, P.M.; Li, Y.; Richardson, G.D.; Vasvari, F.P. Does it really pay to be green? Determinants and consequences of proactive environmental strategies. *J. Account. Public Policy* **2011**, *30*, 122–144. [[CrossRef](#)]
24. Ruggiero, S.; Lehtonen, H. Renewable energy growth and the financial performance of electric utilities: A panel data study. *J. Clean. Prod.* **2017**, *142*, 3676–3688. [[CrossRef](#)]
25. Sueyoshi, T.; Goto, M. Can environmental investment and expenditure enhance financial performance of US electric utility firms under the Clean Air Act amendment of 1990? *Energy Policy* **2009**, *37*, 4819–4826. [[CrossRef](#)]
26. Telle, K. “It pays to be green”—A premature conclusion? *Environ. Resour. Econ.* **2006**, *35*, 195–220. [[CrossRef](#)]
27. Pätäri, S.; Arminen, H.; Tuppura, A.; Jantunen, A. Competitive and responsible? The relationship between corporate social and financial performance in the energy sector. *Renew. Sustain. Energy Rev.* **2014**, *37*, 142–154.
28. Arslan-Ayaydin, Ö.; Thewissen, J. The financial reward for environmental performance in the energy sector. *Energy Environ.* **2016**, *27*, 389–413. [[CrossRef](#)]
29. Sueyoshi, T.; Goto, M. Data envelopment analysis for environmental assessment: Comparison between public and private ownership in petroleum industry. *Eur. J. Oper. Res.* **2012**, *216*, 668–678. [[CrossRef](#)]
30. Doumpos, M.; Andriosopoulos, K.; Galariotis, E.; Makridou, G.; Zopounidis, C. Corporate failure prediction in the European energy sector: A multicriteria approach and the effect of country characteristics. *Eur. J. Oper. Res.* **2017**, *262*, 347–360. [[CrossRef](#)]
31. Bobinaite, V. Financial sustainability of wind electricity sectors in the Baltic States. *Renew. Sustain. Energy Rev.* **2015**, *47*, 794–815. [[CrossRef](#)]
32. Altman, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [[CrossRef](#)]
33. Beaver, W.H. Financial Ratios as Predictors of Failure. *J. Acc. Res.* **1968**, *4*, 71–111. [[CrossRef](#)]
34. Ohlson, J.A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *J. Acc. Res.* **1980**, *18*, 109–131. [[CrossRef](#)]
35. Frydman, H.; Altman, E.I.; Kao, D.L. Introducing recursive partitioning for financial classification: The case of financial distress. *J. Finance* **1985**, *40*, 269–291. [[CrossRef](#)]



36. Koh, H.C.; Kee Low, C. Going concern prediction using data mining techniques. *Manag. Audit. J.* **2004**, *19*, 462–476.
37. Li, H.; Sun, J.; Wu, J. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Syst. Appl.* **2010**, *37*, 5895–5904. [[CrossRef](#)]
38. Gepp, A.; Kumar, K.; Bhattacharya, S. Business failure prediction using decision trees. *J. Forecast.* **2010**, *29*, 536–555. [[CrossRef](#)]
39. Fedorova, E.; Gilenko, E.; Dovzhenko, S. Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Syst. Appl.* **2013**, *40*, 7285–7293. [[CrossRef](#)]
40. Zięba, M.; Tomczak, S.K.; Tomczak, J.M. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Syst. Appl.* **2016**, *58*, 93–101. [[CrossRef](#)]
41. Balcaen, S.; Ooghe, H. 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *Br. Acc. Rev.* **2006**, *38*, 63–93. [[CrossRef](#)]
42. Kim, S.Y.; Upneja, A. Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Econ. Model.* **2014**, *36*, 354–362. [[CrossRef](#)]
43. Larose, D.T. *Discovering Knowledge in Data; An Introduction to Data Mining*; Wiley-Interscience, A John Wiley Sons, Inc. Publication: Hoboken, NJ, USA, 2005.
44. Kass, G.V. An exploratory technique for investigating large quantities of categorical data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1980**, *29*, 119–127. [[CrossRef](#)]
45. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Wadsworth Int. Group: Wadsworth, OH, USA, 1984; Volume 237–251, p. 37.
46. Loh, W.Y.; Shih, Y.S. Split selection methods for classification trees. *Statistica Sinica* **1997**, *7*, 815–840.
47. IBM Knowledge Center. 2019. Available online: [https://www.ibm.com/support/knowledgecenter/SSLVMB\\_subs/statistics\\_casestudies\\_project\\_ddita/spss/tutorials/tree\\_missing\\_crt\\_results.html](https://www.ibm.com/support/knowledgecenter/SSLVMB_subs/statistics_casestudies_project_ddita/spss/tutorials/tree_missing_crt_results.html) (accessed on 15 October 2019).
48. Frank, K. Impact of a Confounding Variable on the Inference of a Regression Coefficient. *Sociol. Methods Res.* **2000**, *29*, 147–194. [[CrossRef](#)]
49. He, G.; Bai, L.; Ren, H.M. Analyst coverage and future stock price crash risk. *J. Appl. Acc. Res.* **2019**, *20*, 63–77. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).