Scientax
Jurnal Kajian Ilmiah Perpajakan Indonesia

djp

VOLUME 4 NO 2 | APRIL 2023 Journal homepage: ejurnal.pajak.go.id ISSN 2686-5718

Application of Data Mining Techniques for VAT-Registered Business Compliance

Yusrifaizal Gumilar Winata[a], Marmah Hadi[b]
[a] Directorate General of Taxes, Jakarta, Indonesia. Email: yusrifaizal@live.com
[b] Polytechnic of State Finance STAN, Jakarta, Indonesia. Email: marmah_hadi@pknstan.ac.id

*Corresponding Author: yusrifaizal@live.com

ABSTRACT

World Bank recommends that Indonesia lower the turnover threshold required to be a VAT-Registered Business from Rp. 4,8 billion to Rp. 600 million to increase VAT-Registered Businesses numbers which will also increase VAT revenue. The number of VAT-Registered Businesses will be significantly increased, which will push Directorate General of Taxes to determine the correct audit priority because it is impossible to audit all taxpayers. This study aims to form a prediction model for formal compliance of VAT-Registered Businesses in the Sampit Tax Office towards 1270 VAT-registered Businesses as of December 31, 2019, which are classified as low-risk VAT-Registered Businesses. The prediction model will be useful for determining audit priorities for certain taxpayers. This study uses a qualitative method using the RapidMiner application and decision tree technique in making prediction models for VAT-Registered Business compliance. The model made has Prediction Efficiency of 67,9%, reduction in Examination Effort by 63.67%, and Strike Rate of 85.99%. The model made is used to predict new VAT-Registered Business data which registered in 2020 and predicts 76 VAT-Registered Businesses will be compliant and 7 VAT-Registered Businesses will not be compliant.

Keywords: data mining, VAT-Registered Business, tax compliance, decision tree

1. INTRODUCTION

Directorate General of Taxes (DGT) has a role in collecting optimal state revenues from taxes. In carrying out its duties, DGT is mandated to contribute to the goals of the Kementerian Keuangan, which are healthy and sustainable fiscal management, optimal state revenues, and agile, effective, & efficient bureaucracy and public services. This is related to one of DGT's missions as stated in KEP-389/PJ/2020 Concerning the Strategic Plan of the Directorate General of Taxes for 2020-2024, namely increasing tax compliance through quality and standardized services, effective education and supervision, and fair law enforcement.

Regarding the mandate to contribute to optimal revenue, judging by the Annual Report of DGT for 2017-2019, the DGT has not succeeded in achieving the revenue target that has been set for the past few years. Considering that state revenues in the State Budget are mostly sourced from taxes, this is a problem that needs to be considered and addressed. According to Prastowo, tax revenues in recent years were not achieved due to several factors, namely the decline in commodity prices, the weakening of the global economy which reduced the realization of import VAT, the number of tax incentives, the use of data and information

doi:10.52869/st.v4i2.317
Received: Received: July 19, 2022; Accepted: March 30, 2023; Published: April 27, 2023

Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia is Sinta 4 Journal (https://sinta.kemdikbud.go.id/journals/profile/9121)
How to Cite:
Winata, Y. G., & Hadi, M. Application of data mining techniques for VAT-Registered Business compliance. *Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia*, *4*(2), 243–260. https://doi.org/10.52869/st.v4i2.317

that had not been optimal, and delays in tax collection of some sectors (Victoria, 2020).

Faced with these factors, the government has taken various actions to address them through IT system upgrades, organizational restructuring, staff skills upgrading, and changes to tax-related policies. Although various actions have been taken to increase tax revenues, according to The World Bank (2020), the actions that have been taken by the Indonesian government have not been sufficient. The actions that have been taken could increase revenues substantially, but the World Bank states that they will take a long time to be fully implemented. The World Bank suggests taking more fundamental steps, such as lowering the threshold for the VAT-Registered Business (PKP) which is currently Rp. 4.8 billion to Rp. 600 million so that more companies can be subject to VAT. Looking at the DGT annual report in 2019, PPN (or VAT) and PPnBM revenues contributed 39.89% (Rp.531.6 trillion) of the total tax revenue of Rp.1.332.66 trillion as shown in figure 1.

Although it has contributed significantly to state revenues, the World Bank believes that the VAT collected by Indonesia has only reached 60% of its original potential (DDTCNews, 2021). With the assumption that the VAT of Rp. 531.6 trillion collected is only 60% of its potential, in 2019 VAT could reach Rp. 886 trillion if it reaches 100% of its potential. This of course can help tax revenue in achieving its target.

At the same time, tax crime activities often occur related to fictitious tax invoices that are closely related to taxable entrepreneurs. This can be seen in the 2019 DGT annual report which continues to investigate criminal acts in the taxation sector, which consists of 79 cases out of a total of 144 case files are related to the VAT sector. If the number of PKP increases rapidly, it is feared that the DGT will be overwhelmed in mitigating the risk of non-compliant PKP. To mitigate this risk, there are many ways that can be taken, one of which is to analyze PKP data using Data Mining techniques.

Data Mining is a process of mining or collecting important information in the form of correlations, patterns, and trends from large data using artificial intelligence, statistical techniques, mathematics, machine learning, and so on (Larose, 2014). Data Mining can be useful for handling large amounts of data and classifying registered PKP compliance as well as determining potential future PKP compliance. With Data Mining, it is expected to be able to mitigate the risk of non-compliant PKP due to the significant PKP registration due to the lowering of the PKP turnover threshold.

The authors picked Sampit Tax Office as research object because Sampit Tax Office has wide area coverage that cover three districts/cities, it has two Tax Services, Dissemination, and

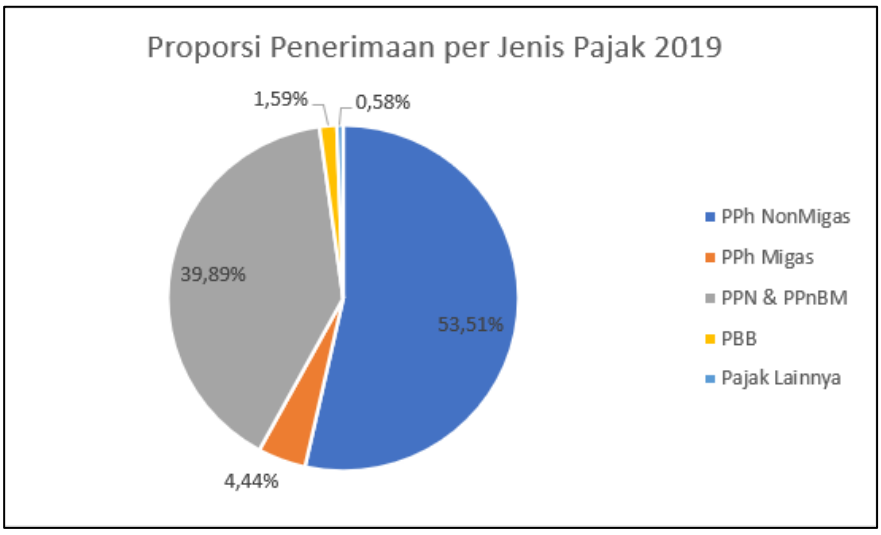


Figure 1 Proportion of Revenue per type of Tax in 2019
Source: Processed from the Annual Report of DGT 2019

Consultation Office (branch offices), and the number of PKP steadily increased every year. Sampit Tax Office also succeeded in revealing a tax crime case related to fictitious tax invoices in 2019 (Norjani, 2019). Hence, Sampit Tax Office is expected to be able to represent other tax offices especially tax offices located outside of Java that usually has wide area coverage and has Tax Services, Dissemination, and Consultation Office.

Reading previous research, Gupta and Nagadevara (2007) state that the use of models through data mining provides a higher level of accuracy and efficiency in tax audits compared to using random selection although sometimes the model produces false negative predictions, namely when taxpayers are actually compliant but classified by the model as non-compliant, and false positive, which is when the actual taxpayer is not compliant but is classified by the model as compliant. Given that the DGT audit coverage ratio has only reached 1.58% according to the 2019 DGT annual report, it is hoped that the use of data mining can help increase the effectiveness of audit activities and produce a deterrent effect on other taxpayers. Supporting that, Wu et al. (2012) explain that data mining will provide more scientific and resource-saving approach compared to manual screening method. Data mining will help tax officials to perform screening more efficiently and enhance the productivity of tax audits.

The use of data mining in the taxation sector has also been studied several times in Indonesia. In the central tax sector, Bagaskara (2018) has applied data mining to create a new individual taxpayer compliance prediction model. Then (Prabowo, 2018) has applied data mining to improve the accuracy of tax socialization. In the local tax sector, Widayati (2018) has applied data mining to create a model for compliance with Land and Building Taxes in the Rural and Urban sectors so that Local Governments can focus on conducting socialization to areas that are less compliant.

Compared to previous researches that conducted research on Taxpayer compliance based on Annual Tax Return and Land and Building Taxes, which also annually, this study conducts research on PKP compliance based on their Monthly VAT report (SPT Masa PPN) compliance. This study would be able to produce a PKP compliance prediction model in order to mitigate the risk of non-compliant PKP if the PKP turnover threshold is lowered.

The model will be able to help Tax Offices especially Sampit Tax Office to allocate its human resources for providing guidance and supervision to PKP which predicted to be Non-Compliant.". In this study, the authors will apply data mining techniques to PKP data provided by Sampit Tax Office, with the title "Application of Data Mining Techniques for VAT-Registered Business Compliance".

2. THEORETICAL FRAMEWORK

2.1 Theory of Planned Behavior

Ajzen (1991) describes the Theory of Planned Behavior (TPB) as a person's intention to perform certain behaviors. A person's intentions can be seen from the attitude towards behavior, subjective norms related to, and control over perceived behavior. A measured effort of an individual to perform a certain behavior will reflect the individual intentions. TPB also argues that a person will perform certain behaviors depending on the benefits and costs of an activity.

In short, a certain behavior can be explained based on a person's intentions. When associated with this research, PKP's intentions should be seen from their behavior. Then the behavior can be analyzed using data mining to make prediction models.

2.2 Theory of Tax Evasion

The difference between tax avoidance and tax evasion lies in the legality of the actions taken by taxpayers. Sandmo (2005) explains tax avoidance as an act of legal tax avoidance by taking

advantage of legal loopholes to reduce tax payables. Taxpayers will not worry if tax avoidance activities are detected by the tax authorities because their actions do not violate the law. Meanwhile, tax evasion is explained as evading taxes through violating the law by not reporting income that should be taxed or carrying out other illegal activities that make him responsible for administrative actions from the tax authorities. Supporting this, Rahmawati and Ibrahim (2015) also differentiate tax avoidance as effort to reduce tax by manipulate within taxpayer own affairs within law, and tax evasion as effort to reduce tax by using illegal manipulation.

In this case, taxpayers will be worried if tax evasion activities were detected by the tax authorities because their actions violate the law. In relation to TPB, if the taxpayer has the intention to carry out a tax evasion, the taxpayer will show a certain behavior. Related to that matter, Valenty and Kusuma (2019) explain that taxpayer behavior to behave compliantly is influenced significantly positive by taxpayer intention to comply.

2.3 Enforcement through the Deterrence Model Theory

Dealing with tax evasion from taxpayers, (Andreoni et al., 1998) have the assumption that taxpayers will not report and pay taxes if there is no enforcement. In relation to this assumption, Wilks and Pacheco (2014) state that tax authorities should focus on preventing tax evasion through tax audits and penalties for non-compliant (deterrence model).

To improve the accuracy of tax audits, Gupta and Nagadevara (2007) have shown that the use of data mining models will provide better strike rate (percentage of taxpayers who are predicted to be non-compliant are taxpayers who are actually non-compliant) compared to the random selection of taxpayer audits which expected to cause a deterrence effect on other taxpayers, so that it will reduce the tax evasion carried out by taxpayers.

2.4 Taxpayer Compliance

Taxpayer compliance according to Rahayu (2020) is divided into two types, namely formal compliance and material compliance. The two types of compliance can be distinguished as follows:

a) Formal compliance, formal compliance is that the taxpayer fulfills his obligations formally in accordance with the provisions of the tax law, with a simple example being reporting his tax return (SPT) on time; and
b) Material compliance, material compliance is that taxpayers substantively fulfill all material provisions of taxation according to the content and spirit of the tax law. An example of material compliance is paying tax payments correctly according to the law.

2.5 VAT-Registered Business (PKP)

VAT-Registered Business (PKP), according to Law Number 8 of 1983 on Value Added Tax of Goods and Services and Sales Tax on Luxury Goods as Already Amended Several Times the Latest by the Decree of Law Number 21 of 2021 , is a person or entity in whatever form within the company or work environment produces goods, imports goods, exports goods, conducts trading business, or conducts service business. However, it excludes small entrepreneurs whose have turnover below Rp. 4.8 billion, which is regulated in Ministry of Finance Regulation Number (No.) PMK-197/PMK.03/2013 Concerning Batasan Pengusaha Kecil Pajak Pertambahan Nilai (Limitations of Small Entrepreneurs for Value Added Tax, unless the small entrepreneur chooses to be registered as a PKP.

In this regard, the World Bank is of the opinion that the turnover limit of Rp. 4.8 billion is too high so that the DGT can only collect VAT of 60% of its original potential. The World Bank proposes to lower the turnover limit to Rp.600 million. If the proposal is followed by the DGT, there will be a rapid increase in the inauguration of PKP which risks monitoring not being carried out optimally. The development of data mining models is needed to help PKP monitoring.

2.6 Low-Risk VAT-Registered Business

the DGT makes a list and criteria of low-risk VAT-registered business who can be given a preliminary refund for tax overpayments. According to Ministry

of Finance Regulation Number (No.) PMK-117/PMK.03/2019 Concerning Pengusaha Kena Pajak Yang Memenuhi Persyaratan Untuk Ditetapkan Sebagai Pengusaha Kena Pajak Berisiko Rendah (VAT-Registered Business Who Meet the Requirements to Be Designated as Low-Risk VAT-Registered Business), low-risk PKP are entrepreneurs who generally meet the following requirements:

a) PKP has submitted the SPT Masa PPN for the last 12 months;
b) PKP is not currently being investigated for preliminary evidence and/or criminal investigations in the taxation sector; and
c) PKP has never been convicted of a crime in the field of taxation based on a court decision that has permanent legal force within the last five years.

These requirements will also be used as criteria in determining compliant and non-compliant PKPs in the formation of the model. Ideally, the research should be conducted based only point B and C since these two points have a high probability of representing taxpayers who actually committed tax fraud. Regarding to that matter, obtaining those samples will be quite difficult since DGT only has audit ratio of 1,58% of the total taxpayers based on Annual Report of Directorate General of Taxes 2019. Besides of that (Vanhoeyveld et al., 2020) explain that the list of PKP that was labelled as non-compliant is a biased representation of the population, since auditors typically audit similar neighboring fraud cases relying on their past experiences.

Because of that shortcoming, this research will tend to focus more on point A. Related to that, Badan Pendidikan dan Pelatihan Keuangan, (2016) and OnlinePajak (2018) point out that PKPs which are non-compliant also tends to not report their VAT Periodic Tax Return. The prediction made by the model will provide early warning for tax officials when the taxpayers register themselves as PKP so tax officials will be able to focus their monitoring to certain new PKP

2.7 Data Mining

Larose and Larose (2014) defines that data mining is the process of extracting data to find a pattern and collect important information from large data. The pattern presented from the data mining must be easy to understand and can be applied to data that will be predicted with a certain degree of certainty. Han et al. (2022) said that data mining is also often referred to as Knowledge Discovery in Databases (KDD) which has process stages consisting of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation. Although data mining itself is actually a part of KDD, Han et al. (2022) explains that KDD is often referred to as data mining because it has a shorter and better known than KDD so that data mining terminology is adopted more broadly as a process of finding interesting patterns and knowledge from large data sources.

Brown (2014) describes that the most popular data mining process, namely CRISP-DM (Cross Industry Standard Process for Data Mining) has six stages of the cycle as follows:

a) Business understanding, namely a clear understanding of the problem to be solved, and how it affects the organization.
b) Data Understanding, namely inspection, explanation, and evaluation of the data held.
c) Data preparation, namely the process of adjusting the data from the data obtained to the form of data prepared for analysis.
d) Modeling, the process of making models with mathematical techniques based on the analyzed data.
e) Evaluation, which is testing whether the resulting model can work well.
f) Deployment, namely integrating the model with the current business.

With this Data Mining technique, it is expected that the research can obtain a model from the data of all PKPs that have been registered in 2019, then the model will be applied to the recently registered PKPs in 2020 to predict compliance so that DGT can allocate resources for supervision more optimally.

2.8 Decision Tree

Rokach and Maimon (2014) describe decision trees in data mining as predictive models that can be used to represent classifiers and regression models

which usually refer to a hierarchical model of decisions and their consequences. Decision making uses a decision tree to identify which options are most likely to achieve a goal.

Kotu and Deshpande (2015) explain decision trees (also referred to as classification trees) as data mining techniques that are easy to prepare and interpret by users, so that decision trees are popular and frequently used as data mining techniques.

Dahan et al. (2014) describe a decision tree consisting of nodes that form a rooted tree which of course has an initial node in the form of a root which then goes through several test nodes or decision nodes by dividing it into a branch which then ends up as a leaf node. Each leaf node represents the target value that is considered the most appropriate based on previous decision making.

3. RESEARCH METHODOLOGY

This research was conducted using a mixed method consisting of quantitative analysis to analyze quantitative attributes (integer and real), and qualitative analysis to analyze categorical attributes (polynomial and binomial). Specifically, this research conducted predictive analysis method that adopts the CRISP-DM which is described by Brown (2014) as a step-by-step data mining process created by data miners for data miners with participants exceeding 200 organizations to create a framework. -his. CRISP-DM applied in this research consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

The objects used in this research are PKP that registered at Sampit Tax Office as per December 2019 as training data and test data for model making. Then the model that has been generated will be applied to the data of the new PKP that registered at the Sampit Tax Office in 2020 as deployment data.

The data used is secondary data in the form of Excel Microsoft Office Open XML Format Spreadsheet file (.xlsx) and comma-separated values (.csv) acquired from Sampit Tax Office with Research Permit Approval numbered S-77/RISET/WPJ.29/ 2021 issued by South and Central Borneo Regional Tax Office on July 2, 2021.

4. RESULT AND DISCUSSION

The research was conducted using supporting applications in the form of Microsoft Excel (MS. Excel) and RapidMiner. The research was conducted through six stages as follows.

4.1 Business Understanding

The business understanding stage is the stage of understanding the business process of how a PKP is classified as compliant and non-compliant based on the low-risk PKP criteria in Ministry of Finance Regulation Number (No.) PMK-117/PMK.03/2019 Concerning Pengusaha Kena Pajak Yang Memenuhi Persyaratan Untuk Ditetapkan Sebagai Pengusaha Kena Pajak Berisiko Rendah (VAT-Registered Business Who Meet the Requirements to Be Designated as Low-Risk VAT-Registered Business . The analysis based on these criteria is as follows:

a) Report SPT Masa PPN (VAT Reports) at least for the last 12 months. In this study, PKP will be classified as compliant if as of December 2019, PKP reports 12 SPT Masa PPN. For PKP which was just registered in 2019, it is considered compliant if reporting the SPT Masa PPN continuously from the time it was registered until the end of the year.
b) PKP will be classified as compliant if the PKP is not recorded in the list of Surat Perintah Pemeriksaan (SP2 or Tax Audit Warrant) for Pemeriksaan Khusus (Special Audit) with audit code 6912, namely a special audit for Taxpayers where there is data and/information indicating taxpayer non-compliant; and
c) PKP will be classified as compliant if the PKP has no tax case recorded in SIPP Pengadilan Negeri Sampit (Sampit District Court's Database) for the last five years.

4.2 Data Understanding

Data requests to Sampit Tax Office are carried out based on the Research Permit Approval numbered

S-77/RISET/WPJ.29/2021 issued by South and Central Borneo Regional Tax Office. The data was acquired by the author in Excel Microsoft Office Open XML Format Spreadsheet file (.xlsx) and comma-separated values (.csv) so that author will use Microsoft Excel in this stage. The author acquired 1353 taxpayer data records that have been registered as PKP at Sampit Tax Office as of 31 December 2020, with financial data for the last four years and 595 SP2 issuance data records published in 2018-2019.

The attributes of the taxpayer data from the data that has been obtained consist of the date of registration, Taxpayer Identfication number, province, city, district, ward, taxpayer status, type of taxpayer, business scope code, legal form, PKP date, PKP number, Tax Return reporting date for the last four years, the turnover of the taxpayer for the last four years, the export of the taxpayer for the last four years, output tax data from transaction partner for the last four years, and the reporting date of the VAT Periodic Tax Return for the last four years.The attributes that will be used in Data Mining then picked based on similar previous researches conducted by Gupta and Nagadevara (2007), Wu et al. (2012), and Bagaskara (2018), despite some attributes from previous researches are unavailable because of confidentiality and time constraint.

4.3 Data Preparation

The attributes of the data acquired by the author need to be prepared to suit the needs of RapidMiner in the data mining process. In this stage, the author uses Ms. excel to add and modify some attributes. Each attribute is described in the following sub-sub-sections.

4.3.1 VAT Compliance Attribute (*Kepatuhan_PPN*)

This attribute will be used as a label for the classification of whether or not the taxpayer complies. This attribute is formed as explained at the business understanding stage with the following results.

From Figure 2 it can be seen that 809 PKPs are classified as compliant (Patuh), and 461 PKPs are classified as non-compliant (Tidak Patuh). However, in determining the classification of processed PKP there are several things that need to be considered, namely:

a) PKPs that do not meet the criteria for Low-Risk PKP based on SP2 issuance data are 5 PKPs based on 2018-2019 inspection data, but these PKPs are no longer registered with PKP as of 2019 so it does not affect the number of PKPs that are classified as non-compliant other than

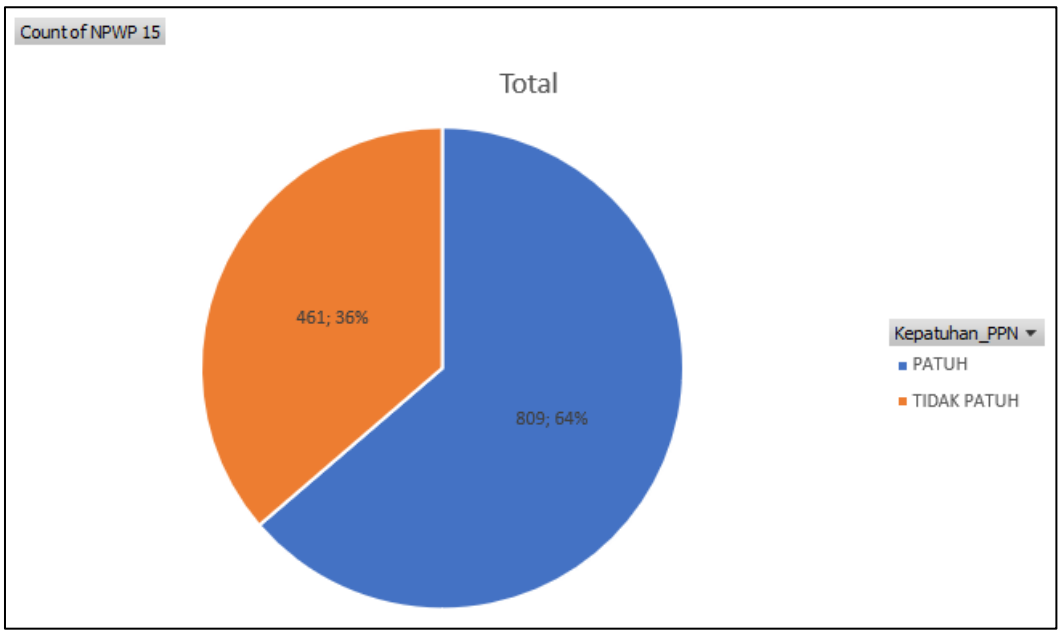


Figure 2 VAT PKP Compliance Classification that has been registered As of December 2019
Source: Processed by Author in 2021

the reporting criteria of VAT Periodic Tax Return, and

b) PKPs that do not meet the criteria for Low-Risk PKP based on SIPP data from the Sampit District Court are 3 PKPs based on data acquired from the Sampit District Court website, one of which is not registered as PKP as of 2019 and the rest do not meet the reporting criteria Periodic Tax Return so that it does not affect the number of PKP which are classified as not complying with the reporting criteria of VAT Periodic Tax Return.

4.3.2 Long Registered Attribute (Lama_terdaftar)

Long Registered attribute is acquired by comparing the difference between the year of Taxpayer Identification Number registration until 2019. Registration of PKP is dominated by taxpayers who have had a Taxpayer Identification Number for 6 years with a total of 103 PKP and as many as 40 taxpayers register themselves as PKP in the same year as the Taxpayer Identification Number registration. In this stage, 1 record of outlier data was found. Data is considered outlier data because it has a value exceeding 100 years.

4.3.3 Tax Administration Office Attribute (Jenis_Kantor)

The jenis_kantor attribute acquired from converting the attributes of the taxpayer city with the value of "Kabupaten Kotawaringin Timur" as Sampit Tax Office, "Kabupaten Katingan" as Kasongan Tax Services, Dissemination, and Consultation Office, and "Kabupaten Seruyan" as Kuala Pembuang Tax Services, Dissemination, and Consultation Office. The conversion results show that there are 891 taxpayers administered at Sampit Tax Office, 235 taxpayers administered at Kasongan Tax Services, Dissemination, and Consultation Office, and 144 taxpayers administered at Kuala Pembuang Tax Services, Dissemination, and Consultation Office.

4.3.4 District Attribute (Kecamatan)

District attribute is obtained from data acquired from the Sampit Tax Office. The author chooses to use the kecamatan attribute instead of the kelurahan (ward) attribute because the kecamatan attribute has fewer classifications than the kelurahan attribute which is expected to simplify the decision tree model if the location of the taxpayer affects VAT compliance.

4.3.5 Distance Attribute (Jarak)

The jarak attribute is obtained from the kecamatan attribute which is transformed into the jarak attribute by using the shortest road route from the taxpayer's place of business to the nearest tax administration office (KPP) according to google maps or using a radius if it is located in the same district with the tax administration office. The distance is then converted into 5 categories, which are 0 (0-30 KM), 1 (31-70 KM), 2 (71-120 KM), 3 (121-180 KM) and 4 (over 180 KM).

4.3.6 Taxpayer Type Attribute (Jenis_WP)

Taxpayer type attribute is not converted because it is already available from the data acquired. The data consists of two types, namely corporate taxpayers (WP Badan) and individual taxpayers (WP Orang Pribadi). The data acquired consisted of 1218 corporate taxpayers and 52 individual taxpayers.

4.3.7 Legal Form Attribute (Bentuk_hukum)

Legal form attribute is not converted because it is already available from the data acquired. The data consists of seven types, namely individuals, BUMN/BUMD, Foundations, Cooperatives, Other Organizations, Partnership and Companies. The distribution of data consists of 52 individuals, 2 BUMN/BUMD, 2 foundations, 72 cooperatives, 3 other organizations, 853 Partnership, and 286 Companies.

4.3.8 Business Scope Code Attribute (KLU)

Business scope code attribute was not converted because it was already available from the data obtained. The types of KLU are regulated in KEP-321/PJ/2012 Concerning Amendments to the

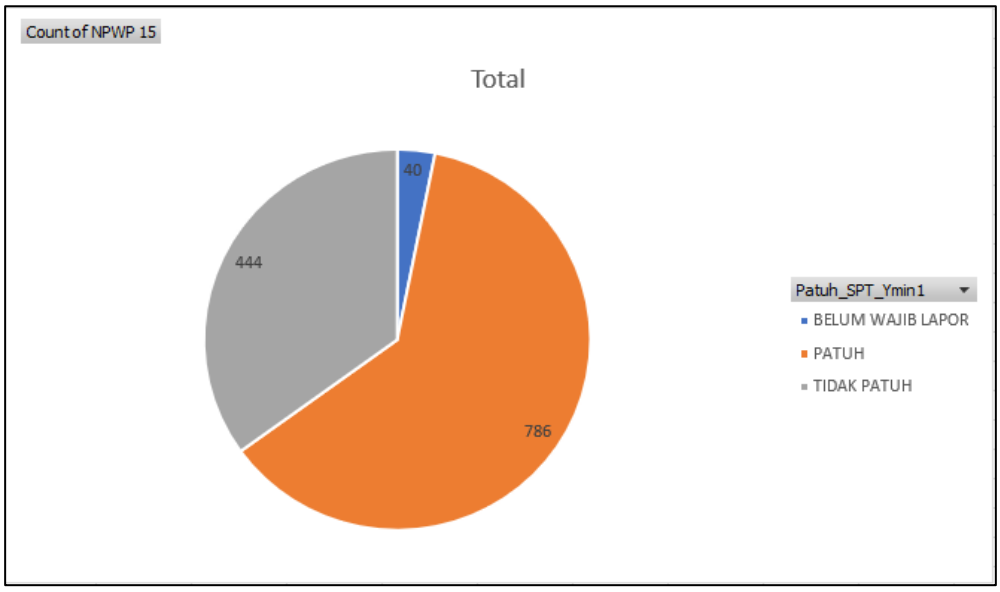


Figure 3 Distribution of previous year's Tax Return Compliance
Source: Processed by Author in 2021

Decree of the Directorate General of Taxes Number KEP-233/PJ/2012 Concerning Classification of Taxpayer Business Fields (Klasifikasi Lapangan Usaha Wajib Pajak). The KLU attribute from the data obtained by the author is dominated by KLU code 46100, namely wholesale trade on the basis of fees or contracts totaling 418 PKP, KLU code 42111, namely road construction with total of 100 PKP, and KLU code 41012, namely office building construction with total of 58 PKP. In this attribute, the writer finds 5 outlier data. The data is considered an outlier because it has the values 'ERR01' and 'ERR04' which are not KLU codes.

4.3.9 Tax Return Compliance Attribute (Kepatuhan Tahunan)

The Tax Return Compliance attribute is an attribute that determines if PKP is compliant based on the previous year's Tax Return which was reported in the year the taxpayer was registered as PKP. If PKP is not registered as Taxpayer on that year, then PKP

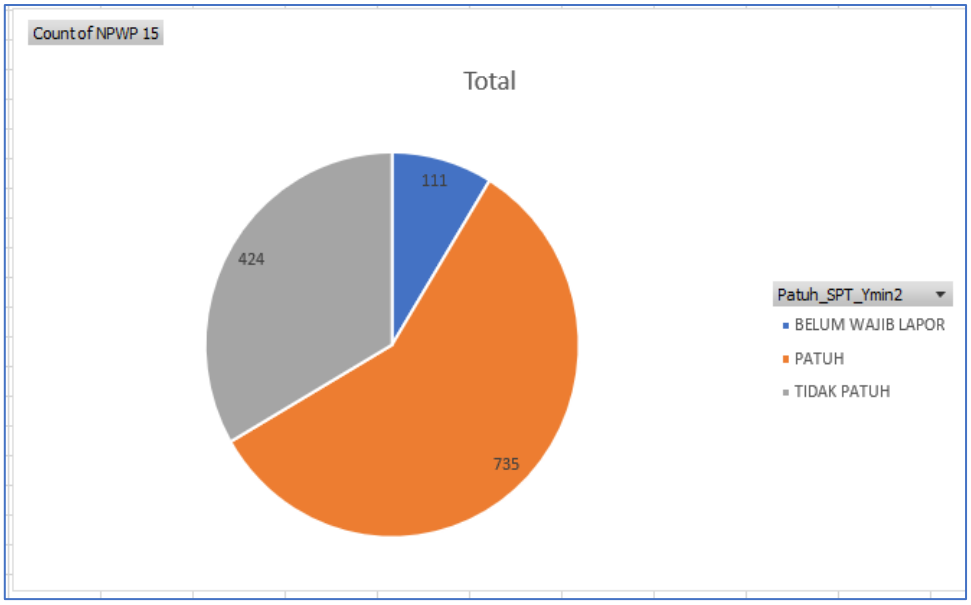
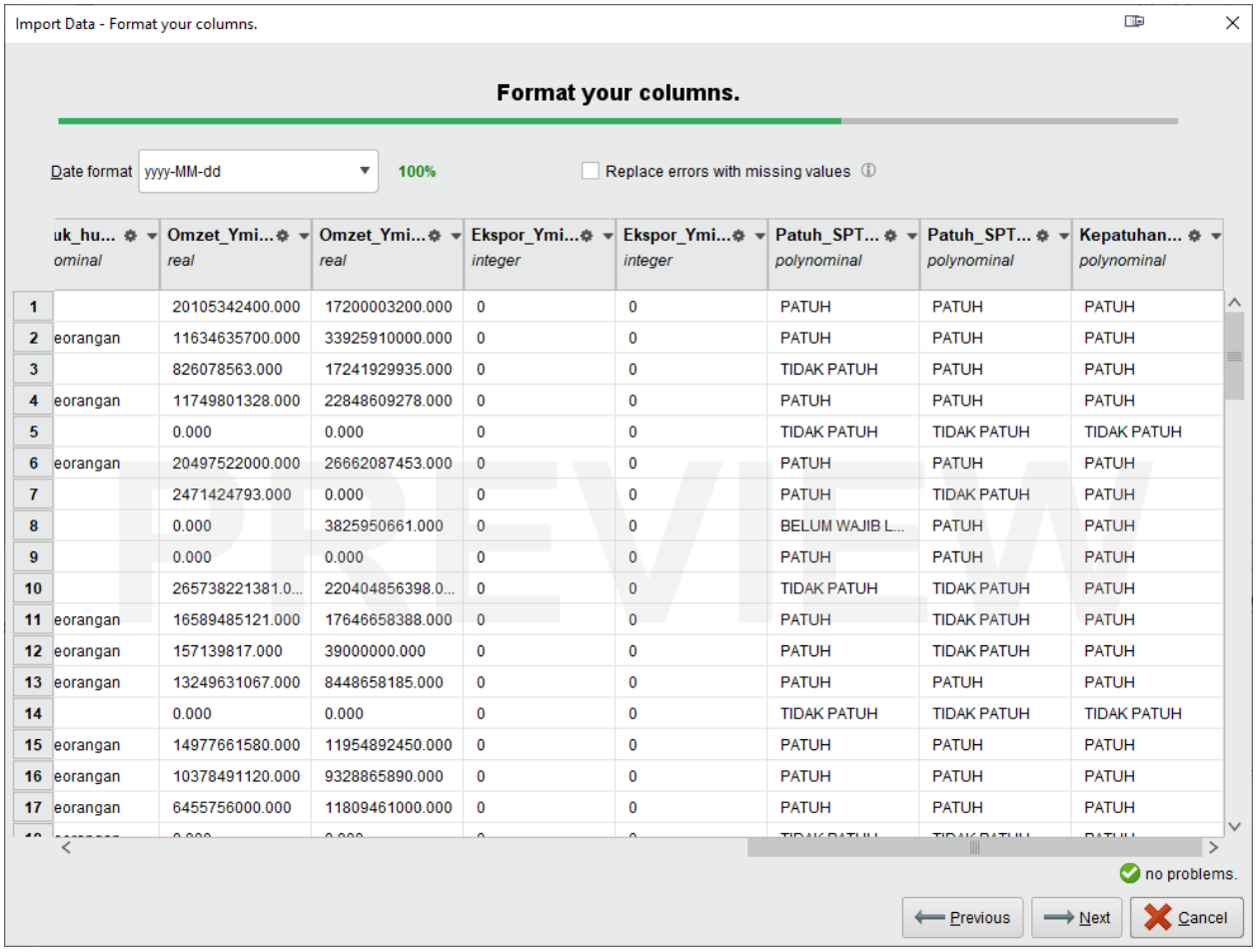


Figure 4 Distribution of Tax Return Compliance Two Years Before
Source: Processed by Author in 2021



	uk_hu... ominal	Omzet_Ymi... real	Omzet_Ymi... real	Ekspor_Ymi... integer	Ekspor_Ymi... integer	Patuh_SPT... polynominal	Patuh_SPT... polynominal	Kepatuhan... polynominal
1		20105342400.000	17200003200.000	0	0	PATUH	PATUH	PATUH
2	eorangan	11634635700.000	33925910000.000	0	0	PATUH	PATUH	PATUH
3		826078563.000	17241929935.000	0	0	TIDAK PATUH	PATUH	PATUH
4	eorangan	11749801328.000	22848609278.000	0	0	PATUH	PATUH	PATUH
5		0.000	0.000	0	0	TIDAK PATUH	TIDAK PATUH	TIDAK PATUH
6	eorangan	20497522000.000	26662087453.000	0	0	PATUH	PATUH	PATUH
7		2471424793.000	0.000	0	0	PATUH	TIDAK PATUH	PATUH
8		0.000	3825950661.000	0	0	BELUM WAJIB L...	PATUH	PATUH
9		0.000	0.000	0	0	PATUH	PATUH	PATUH
10		265738221381.0...	220404856398.0...	0	0	TIDAK PATUH	TIDAK PATUH	PATUH
11	eorangan	16589485121.000	17646658388.000	0	0	PATUH	TIDAK PATUH	PATUH
12	eorangan	157139817.000	39000000.000	0	0	PATUH	TIDAK PATUH	PATUH
13	eorangan	13249631067.000	8448658185.000	0	0	PATUH	PATUH	PATUH
14		0.000	0.000	0	0	TIDAK PATUH	TIDAK PATUH	TIDAK PATUH
15	eorangan	14977661580.000	11954892450.000	0	0	PATUH	PATUH	PATUH
16	eorangan	10378491120.000	9328865890.000	0	0	PATUH	PATUH	PATUH
17	eorangan	6455756000.000	11809461000.000	0	0	PATUH	PATUH	PATUH

Figure 5 RapidMiner Import Process
Source: Processed by Author in 2021

is classified as "Has no obligation to report" (Belum Wajib Lapor). If PKP is registered on that year and report Tax Return before the deadline then PKP is classified as "Compliant" (Patuh). If PKP repot Tax Return late, then PKP is classified as "Not Compliant" (Tidak Patuh). This attribute is further divided into two more attributes, namely "patuh_SPT_ymin1" attribute (based on previous year) patuh_SPT_ymin2 attribute (based on two years before). The distribution of data on the patuh_spt_ymin1 attribute is as follows.

Figure 3 shows that based on previous year's Tax Return Compliance, 786 PKP classified as compliant (patuh), 444 PKP classified as not compliant (tidak patuh), and 40 PKP classified as has no obligation to report (belum wajib lapor). Then for the patuh_spt_ymin2 attribute it has the data distribution as following figure.

Figure 4 shows that based on previous year's Tax Return Compliance, 735 PKP classified as compliant (patuh), 424 PKP classified as not compliant (tidak patuh), and 111 PKP classified as has no obligation to report (belum wajib lapor).

4.3.10 Turnover and Export Attribute (Omzet & Ekspor)

The turnover (omzet) and export (ekspor) attributes were not converted because they were already available from the data acquired. The turnover attribute is divided into two attributes, namely the turnover attribute of the previous year (omzet_ymin1) and the turnover attribute of the previous two years (omzet_ymin2). Furthermore, the export attributes are also divided into two attributes, namely the export attribute of the previous year (ekspor_ymin2) and the export attribute of the previous two years (ekspor_ymin2).

4.3.11 Attribute Deletion

In this stage, the attributes that are considered unnecessary in the data mining process are

removed. Some of the deleted data such as the taxpayer's name, address, province, taxpayer status, business field code name, PKP number, recap of VAT transaction data from VAT Partners (because the data used in deployment does not have that data), and several other attributes are deleted so that the existing data on training data are only data that has been described in the previous stages.

4.4 Modelling

This stage includes extracting data for model making. Data that has been processed using the ms.excel application is exported in csv form which is then imported by the RapidMiner application.

4.4.1 RapidMiner Import Process

The process of importing csv data from the ms.excel application to the RapidMiner application went smoothly. Every column of csv is detected by RapidMiner and there are no empty rows. The details of the import process are as follows.

Figure 5 shows that some data are not classified according to the type of data the author wants. Some of the data that are not classified correctly are the attributes of KLU (integer), tipe_wp (polynominal), Ekspor_Ymin1(integer), Ekspor_Ymin2 (integer), and Kepatuhan_PPN (polynominal). Then the author modifies these attributes with the results of the attribute data type as shown in table 1.

4.4.2 Data Mining

After the data has been successfully imported, it is necessary to filter the outlier data found at the data preparation stage (5 KLU error code data and 1 old registered data > 100 years) using the 'filter examples' operator and set the 'kepatuhan_ppn' attribute as a label, as a prerequisite for use decision tree technique, using the 'set role' operator. Filters are applied as follows.

As discussed in the data preparation stage, in Figure 6, filtering is carried out with the criteria 'kode_klu' is considered normal if it does not contain the word 'ERR' and 'lama_daftar' is considered normal if it is less than 100 years. After that, the 'set_role' operator is applied with a picture as shown in the screenshot in Figure 6.

After the data is adjusted, it is necessary to choose the right prediction model. Han et al. (2022) explained that the Receiver Operating Characteristic Curve (ROC Curve) is useful visual tools for comparing modeling techniques in order to evaluate each model and find the best model. Then the author will test the ROC Curve to ensure that the decision tree technique can be used and applied better than other techniques. The ROC Curve test was carried out by comparing the decision tree, random forest, k-NN, and deep learning with the following results.

Table 1 Description of Data Types in the RapidMiner Application
Sumber: Processed by Author in 2021

Attribute	Data type
Lama_daftar (Long Registered)	integer
jenis_kantor (Tax Administration Office)	polynominal
Jarak (Distance)	integer
Kecamatan (District)	polynominal
jenis_wp (Taxpayer type)	binominal
kode_klu (Business Scope)	polynominal
bentuk_hukum (Legal Form)	polynominal
omzet_ymin2 (turnover 2 years before)	real
omzet_ymin1 (turnover in prev year)	real
ekspor_ymin2 (export 2 years before)	real
ekspor_ymin1 (export in prev year)	real
patuh_spt_ymin2 (tax return 2 yrs before)	polynominal
patuh_spt_ymin1 (tax return in prev year)	polynominal
kepatuhan_ppn (VAT Compliance)	binominal

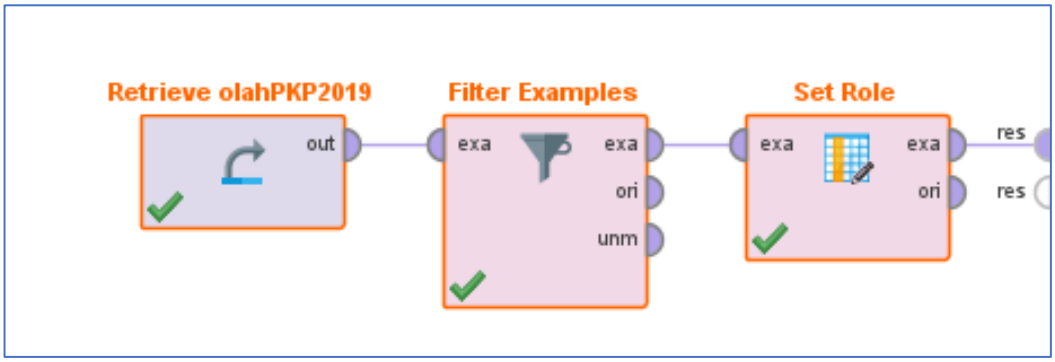


Figure 6 Data Adjustment Process before Data Mining
Source: Processed by Author in 2021

The ROC Curve uses the Y axis as the True Positive Rate (non-compliant PKP is classified as non-compliant PKP) and the X axis as the False Positive Rate (compliant PKP is classified as non-compliant PKP). Radečić (2020) explains that a good model has a True Positive Rate exceeding 0.5 and it will be better if it is closer to 1 (the graph will tend to be in the upper left zone) because if the True Positive Rate is 0.5 then it is not better than guessing randomly and if it is less than 0.5 it means that the prediction made by the model is not correct. In Figure 7, the four selected models exceed True Positive Rate 0.5, but the decision tree represented by the green line is the best model because it is closest to True Positive Rate 1 and the overall graph distribution is in the farthest upper left position compared to other models.

4.4.3 Application of Decision Tree

At this stage, the decision tree is applied using an additional operator 'cross validation' in order to provide performance test results. The selected Cross Validation has a number of folds by default of 10 which means that it will divide the data into 10 smaller data sections. Then the cross validation operator will use 9 parts for training data, and 1 part for test data. The RapidMiner application will

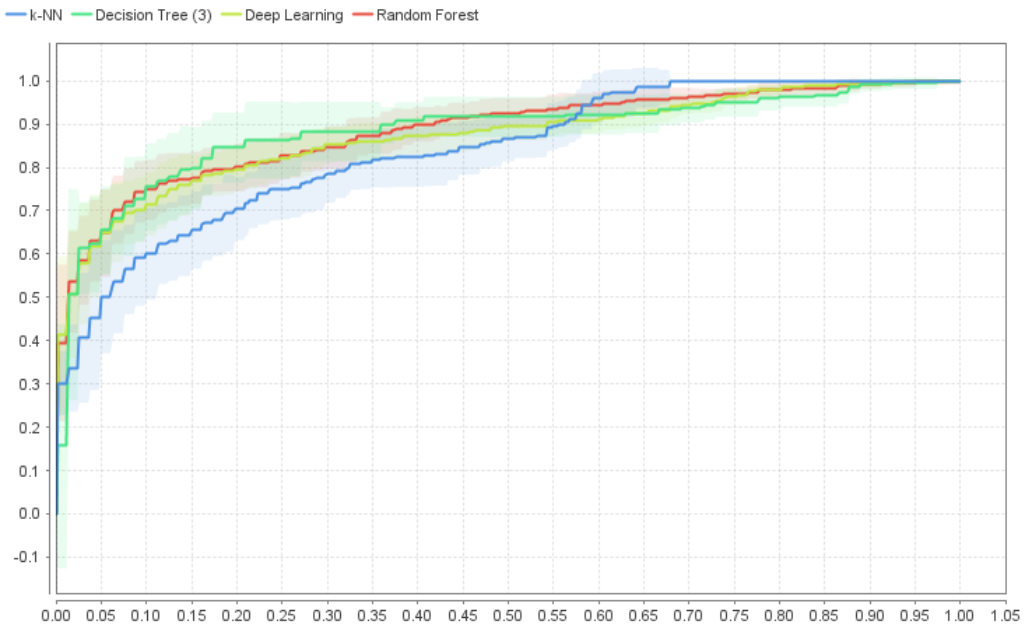


Figure 7 ROC Curve test result
Source: Processed by Author in 2021

accuracy: 81.24% +/- 3.57% (micro average: 81.25%)

	true PATUH	true TIDAK PATUH	class precision
pred. PATUH	740	170	81.32%
pred. TIDAK PATUH	68	291	81.06%
class recall	91.58%	63.12%	

Figure 8 The Results of Implementing the Decision Tree
Source: Processed by Author in 2021

conduct training and testing 10 times until each part has been tested. Next, the decision tree operator uses the 'gain_ratio' criterion (default) with the following results.

Based on Figure 8, the model formed has an accuracy rate of 81.24% with the ability to accurately predict 'compliant' correctly 81.32% of all subjects who were predicted 'compliant' and predict "non-compliant" correctly 81.06%. of all subjects who were predicted to be 'non-compliant. Then the model has a True Positive Rate or the ability to correctly classify PKP 'compliant' at 91.58% and True Negative Rate or the ability to classify PKP as 'non-compliant' correctly at 63.12%.

The results of this test need to be known to find out how the model will perform for now and help comparing models later. In order to improve the accuracy of the model, author apply "weights by information", "forward selection", and "Optimize Parameter".

As seen in Figure 9, the decision tree model will be optimized using weight by information (weight), forward selection (FS), weight+FS, or Optimize Parameters and then they will be tested

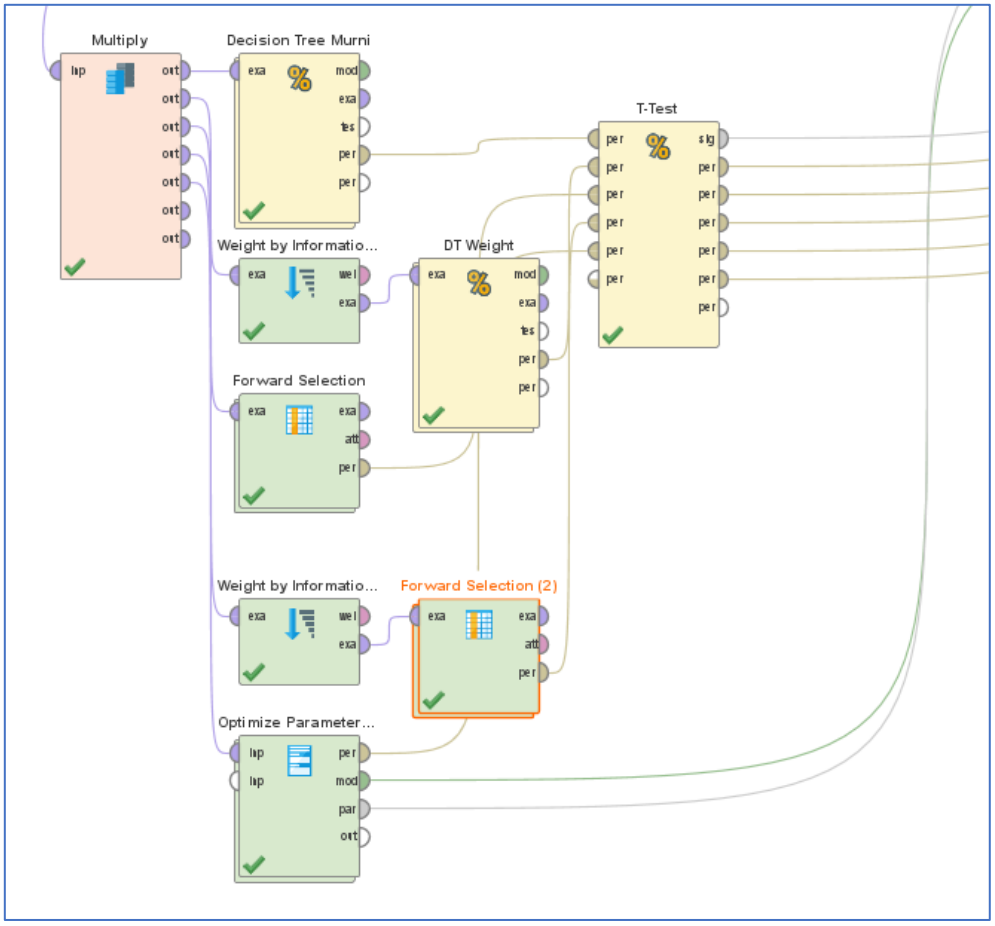


Figure 9 Improving Models Accuracy
Source: Processed by Author in 2021on the RapidMiner application

Table 2 Comparison of Accuracy Among Models
Sumber: Processed by Author in 2021

Model	DT	Weight	FS	Weight +FS	Optimize Par
Accuracy	81,24%	83,69%	84,24%	84,16%	84,32%

Table 3 Confusion Matrix of Selected Model
Sumber: Processed by Author in 2021

	Actual Positive	Actual Negative	Class Precision
Predicted Positive	757 (TP)	148 (FP)	83,65%
Predicted Negative	51 (FN)	313 (TN)	85,99%
Class Recall	93,69%	67,90%	

Table 4 Performance Comparison Between Model
Sumber: processed by the author using RapidMiner

Model	PE	EF	SR
DT	63,12%	63,67%.	81,06%
Weight	67,46%	63,67%.	83,16%
FS	67,46%	63,67%.	86,15%
Weight+FS	68,11%	63,67%.	85,33%
Optimized Par	67,90%	63,67%.	85,99%

using a T- test to compare the accuracy Based on table 2, it can be seen that the use of a decision tree model that is optimized using Optimize Parameter has a higher accuracy rate compared to other optimization methods and has a simpler model. Therefore, decision tree model that uses Optimize Parameters will be used for data deployment. levels among models with the following results.

The Optimize Operator Parameters used have the following information:

a) Using criterion parameter consisting of gain_ratio and information_gain because the data has several attributes with polynomial and binominal data types; and
b) Using the minimum_gain parameter to control the split in the decision tree with a minimum value of 0.01, a maximum value of 1, and 100 steps.

4.5 Evaluation

The model selected in the previous stage has a confusion matrix as table 3.

According to research conducted by Gupta and Nagadevara (2007), table 3 can be evaluated as follows:

a) The model made has the ability to correctly predict non-compliant PKP from all non-compliant PKP (Prediction Efficiency), or Class Recall in the table 3, with the formula PE = TP/ (TP+FN) or 313/313+148 = 67.90%.
b) The model created can reduce the effort required in the audit if every taxpayer is audited (reduction in Examination Effort) with the formula
c) EF = 1- (FP + TN)/ Total cases or 1- (148+313)/1269 = 63, 67%.
d) The model made has the ability to get PKP that are truly non-compliant if all PKPs that are predicted to be non-compliant are audited (Strike Rate), or Class Precision in the table 3, with the formula SR= TP/(TP+FP) or 313/313+51 = 85.99%.

Comparison of application with previous models can be seen in table 4.

In comparing between models, the strike rate is preferred because the main purpose of modeling is to be able to employ resources more

effectively. But that does not mean that prediction efficiency can be ignored, because if the prediction efficiency is low, then it will be no better than random selection. Based on table 4, the selected model, which is optimized parameter, has a higher strike rate than the other models. Then tree from the model can be described in Figure 10.

Figure 10 shows the tree of the decision tree of the selected model. The selected root node, which is the 'Patuh_SPT_Ymin1' attribute, is checking whether the taxpayer has reported the previous year's Annual Tax Return (SPT). Overall leaf node of the decision tree ends at 'Turnover_Ymin2', namely PKP turnover in the previous two years, lama_daftar (long registered), and Jenis_kantor (Tax administration Office type). In general, if the PKP has reported the previous year's Annual Tax Return on time, then the PKP tends to be classified as a compliant PKP or a low risk PKP. The next attribute that affects the compliance is the turnover attribute. If the previous year's turnover was between Rp. 2.6 - 24 billion, the Taxpayer Identification Number has been.

Registered for more than 10 years and has a distance category exceeding 1, then the PKP will tend to be classified as non-compliant. If the PKP has registered Taxpayer Identification Number for more than 1 year, has a turnover of more than Rp. 37 million, and the distance category exceeds 2, it will be classified as non-compliant. If the PKP has a distance category of less than 3, and the turnover in the previous two years exceeds Rp. 1.5 billion, it tends to be classified as non-compliant. If the PKP has a turnover in the previous 2 years of less than Rp. 1.5 billion and is registered at an administrative office other than the Tax Office, it will be classified as non-compliant.

The implication of this model is that Sampit Tax Office needs to provide guidance to PKPs who did not comply with the previous year's Annual Tax Return (SPT) because if their annual formal obligations are not fulfilled, then their monthly formal obligations are also at risk of not being fulfilled. More specifically, Sampit Tax Office needs to supervise and educate PKP which are within the administrative area of its Tax Services, Dissemination, and Consultation Office, PKP which

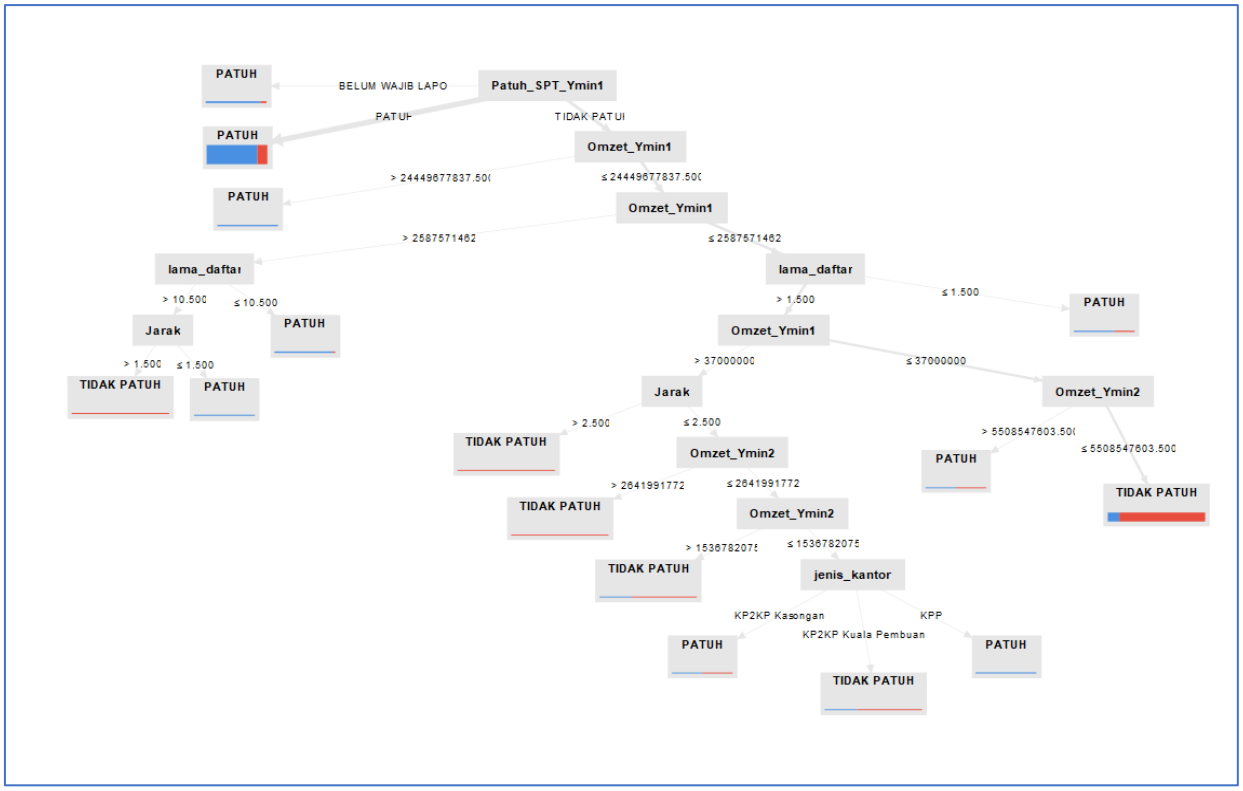


Figure 10 Decision Tree of Selected Model
Source: Processed by Author in 2021 using RapidMiner

Table 5 VAT Compliance Classification Results on Data Deployment
Sumber: Processed by Author in 2021using RapidMiner

Classification	Amount of PKP	*Max Confidence*	*Average Confidence*
Compliant	76	0,9	0,791
Non-Compliant	7	1	0,209

is in the category of distance exceeding 2 (more than 61 KM). Then the model also implies that PKPs with a turnover of less than the small business threshold (Rp.4.8 billion) tend to be non-compliant, so the Sampit Tax Office needs to provide more guidance to PKPs who choose themselves to be registered as PKP than PKPs who are registered in office by Tax Office.

4.6 Evaluation

After the classification model is made, the model is then applied to new data as data deployment in the form of taxpayer data that has just been registered as PKP in 2020 at Sampit Tax Office in order to predict the behavior of the newly registered PKP. The results of the application of the model to the deployment of data obtained are as follows. Based on table 5, it can be seen that 76 PKPs are predicted to become compliant and 7 PKPs are predicted to be non-compliant. The compliant classification has a maximum confidence of 0.9 and the non-compliant classification has a maximum confidence of 1. The average confidence of the non-compliant classification is much lower because there are fewer PKPs classified as non-compliant. PKPs classified as non-compliant come from PKPs who have had a Taxpayer Identification Number for a long time but do not report their Annual Tax Return on time and are registered as PKP at their own request, seen from their turnover, which is less than Rp. 4.8 billion in a year.

5. CONCLUSIONS

Based on the research described above, the conclusion can be drawn is that by using historical data on tax reporting in 2017-2019 at Sampit Tax Office, a Receiver Operating Characteristic Curve (ROC Curve) test was carried out to compare better modeling techniques between Decision Tree models which are Random Forest, k-NN, and Deep Learning. The ROC Curve test results show that the Decision Tree model is the best model compared to the other three models.

Modeling with a decision tree is carried out using historical data on taxpayer compliance in 2017-2019 as training data and test data to test PKP compliance in fulfilling VAT Periodic Tax return (SPT Masa PPN). This model can classify taxpayer non-compliance correctly from all non-compliant taxpayers (Prediction Efficiency / PE) of 67.9%. Then the model can contribute to reducing the number of examinations compared to random selection (reduction in Examination Effort / EF) by 63.67%. The level of accuracy of this model in predicting non-compliant PKP correctly if every predicted PKP is audited (Strike Rate / SR) is 85.99%.

Based on the model that has been made, the attributes that determine taxpayer compliance in the VAT Reporting compliance are Annual Tax Return compliance (SPT Tahunan), business turnover, length of time the taxpayer was registered, distance from the nearest tax administration office, and tax administration area (Tax Office or Tax Services, Dissemination, and Consultation Office).

Furthermore, the model created from the training data is applied to the PKP which was just registered in 2020 as a data deployment resulting in predictions of 76 Compliant Taxpayers and 7 Non-Compliant Taxpayers.

6. IMPLICATIONS AND LIMITATIONS

The model that has been made can be used by Sampit Tax Office to predict whether or not taxpayers who have just been registered as PKP and should be followed up by providing more guidance and supervision to PKPs who have had a Taxpayer Identification Number for a long time but

do not report their Annual Tax Return (SPT Tahunan) on time and are registered as PKP by their own request.

In the preparation of this research, there are still some shortcomings and limitations, such as the limited time and data used in the study which was carried out for less than one semester so that the author could not use more data and methods to be applied. This study also limits the data sample which includes PKP that has been confirmed at Sampit Tax Office as of December 31, 2020, with financial data and SPT compliance for 2017-2020. In addition, the attributes used are also limited due to confidential information, access restrictions, and databases that are less integrated with each other so that it requires more effort and time to obtain more attributes.

Further research should use wider data samples or conduct research in tax administrative offices that have larger number of PKPs with higher audit coverage ratios and use more attributes to determine whether or not PKP complies based on data on preliminary evidence examinations and/or tax investigations and criminal cases taxation.

REFERENCES

[1] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

[2] Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature - JSTOR*. https://www.jstor.org/stable/2565123

[3] Norjani (2019, December 10). Dua tersangka pengemplang pajak diserahkan ke Kejari Kotim. AntaraNews. https://kalteng.antaranews.com/berita/358871/dua-tersangka-pengemplang-pajak-diserahkan-ke-kejari-kotim

[4] Badan Pendidikan dan Pelatihan Keuangan. (2016, November 9). Kasus Faktur Pajak Fiktif dan Pencegahannya. Badan Pendidikan Dan Pelatihan Keuangan. https://bppk.kemenkeu.go.id/content/berita/sekretariat-badan-kasus-faktur-pajak-fiktif-dan-pencegahannya-2019-11-05-30a15ffe/

[5] Bagaskara, D. (2018). *Penerapan metode data mining untuk klasifikasi kepatuhan pembayaran pajak wajib pajak orang pribadi baru*. Skripsi. Diploma IV Akuntansi (TB). Politeknik Keuangan Negara STAN. .

[5] Brown, M. S. (2014). *Data mining for dummies*. John Wiley & Sons, Inc.

[6] Dahan, H., Cohen, S., Rokach, L., & Maimon, O. (2014). Proactive data mining using decision trees. 21–33. https://doi.org/10.1007/978-1-4939-0539-3_3

[7] DDTCNews. (2021, March 15). Batasan Pengusaha Kena Pajak Bakal Diturunkan. DDTCNews. https://news.ddtc.co.id/batasan-pengusaha-kena-pajak-bakal-diturunkan-28415

[8] KEP-321/PJ/2012 concerning Amendments to the Decree of the Directorate General of Taxes Number KEP-233/PJ/2012 concerning Classification of Taxpayer Business Fields (Klasifikasi Lapangan Usaha Wajib Pajak) (2012).

[9] KEP-389/PJ/2020 concerning the Strategic Plan of the Directorate General of Taxes for 2020-2024, Directorate General of Taxes (2020). https://pajak.go.id/sites/default/files/2020-09/KEP-389PJ2020.pdf

[10] Directorate General of Taxes. (2019). *Continuous Development Of Organizational Capacity through Strengthening the Governance of Taxation Data and Information Technology*. Annual Report 2019.

[11] Law Number 8 of 1983 on Value Added Tax of Goods and Services and Sales Tax on Luxury Goods as already amended several times the latest by the decree of Law Number 21 of 2021, (2021).

[12] Gupta, M., & Nagadevara, V. (2007). Audit selection strategy for improving tax compliance: application of data mining techniques. Foundations of Risk-Based Audits. *Proceedings of the Eleventh International Conference on e-Governance*, 28–30.

[13] Han, J., Pei, J. (Computer scientist), & Tong, H. (2022). *Data mining concepts and techniques (4th ed.)*. Morgan Kaufmann.

[14] Victoria, A.O. (2020, January 8). Pengamat: Target Pajak Tak Pernah Tercapai dalam 10 Tahun Terakhir. Katadata.Co.Id. https://katadata.co.id/agustiyanti/finansial/5e9a4c3b2d85b/pengamat-target-pajak-tak-pernah-tercapai-dalam-10-tahun-terakhir

[15] Kotu, V., & Deshpande, B. (2015). Predictive analytics and data mining: Concepts and practice with rapidminer. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*, 1–425. https://doi.org/10.1016/C2014-0-00329-2

[16] Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining (Vol. 4)*. John Wiley & Sons.

https://books.google.com/books?hl=id&lr=&id=9hOpAwAAQBAJ&oi=fnd&pg=PR11&dq=Discovering+Knowledge+in+Data+:+An+Introduction+to+Data+Mining+(Second+Edition&ots=9Q9y6PeRZ8&sig=nDXm5Gx7jI20Qywgg-tmcAOrZ7o

[17] Ministry of Finance Regulation number (No.) PMK-197/PMK.03/2013 concerning Batasan Pengusaha Kecil Pajak Pertambahan Nilai (Limitations of Small Entrepreneurs for Value Added Tax), (2013).

[18] Ministry of Finance Regulation Number (No.) PMK-117/PMK.03/2019 concerning Pengusaha Kena Pajak yang memenuhi persyaratan untuk ditetapkan sebagai Pengusaha Kena Pajak Berisiko Rendah (VAT-Registered Business who meet the requirements to be designated as Low-Risk VAT-Registered Business), (2019).

[19] OnlinePajak. (2018, December 3). Modus Penerbitan Faktur Pajak Fiktif. OnlinePajak. https://www.online-pajak.com/tentang-ppn-efaktur/modus-penerbitan-faktur-pajak-fiktif

[20] Prabowo, R. A. (2018). *Analisis data call center menggunakan teknik data mining untuk mendukung program sosialisasi perpajakan*. Skripsi. Diploma IV Akuntansi (TB). Politeknik Keuangan Negara STAN.

[21] Radečić, D. (2020, December 8). ROC and AUC — How to Evaluate Machine Learning Models in No Time | by Dario Radečić | Towards Data Science. Towardsdatascience. https://towardsdatascience.com/roc-and-auc-how-to-evaluate-machine-learning-models-in-no-time-fb2304c83a7f

[22] Rahayu, S. (2020). *Perpajakan: Konsep sistem dan implementasi*. http://digilib.itbwigalumajang.ac.id/index.php?p=show_detail&id=18738

[23] Rahmawati, & Ibrahim, S. I. (2015). An appraisal of the tax evasion and tax avoidance system in indonesia. *Akuntabilitas*, *8*(2), 121–132. https://doi.org/10.15408/AKT.V8I2.2880

[24] Rokach, L., & Maimon, O. (2014). Data Mining with Decision Trees: Theory and Applications, 2nd Edition. Data Mining with Decision Trees: Theory and Applications, 2nd Edition, 81, 1–305. https://doi.org/10.1142/9097

[25] Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, *58*(4), 643–663. https://doi.org/10.17310/NTJ.2005.4.02

[26] The World Bank. (2020, July). Indonesia Economic Prospects The Long Road To Recovery. WorldBank.Org, 55–56. https://documents1.worldbank.org/curated/en/804791594826869284/pdf/Indonesia-Economic-Prospects-The-Long-Road-to-Recovery.pdf

[27] Valenty, Y. A., & Kusuma, H. (2019). Determinan kepatuhan pajak: perspektif theory of planned behavior dan teori institusional. *Proceeding of National Conference on Accounting & Finance*, *1*, 47–56. https://doi.org/10.20885/ncaf.vol1.art5

[28] Vanhoeyveld, J., Martens, D., & Peeters, B. (2020). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing*, *86*, 105895. https://doi.org/10.1016/J.ASOC.2019.105895

[29] Widayati, Q. (2018). Penerapan data mining menggunakan metode teknik classification untuk melihat potensi kepatuhan wajib pajak bumi dan bangunan. *Jurnal Ilmiah Matrik*, *20*(2), 157–168. https://doi.org/10.33557/JURNALMATRIK.V20I2.119

[30] Wilks, D. C., & Pacheco, L. (2014). *Tax compliance, corruption and deterrence: An application of the slippery model*. Universidade Portucalense.

[31] Wu, R.-S., Ou, C. S., Lin, H., Chang, S.-I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, *39*(10), 8769–8777.