

FTIR OF HALAL AND NON-HALAL ADULTERATIONS PREDICTION ANALYTICS USING INDUCTION DECISION TREE (ID3) ALGORITHM

AKBAR ARIEF ^{1*}, MIRA KARTIWI ¹, IRWANDI JASWIR ²

¹Kulliyah of Information and Communication Technology, International Islamic University
Kuala Lumpur, Malaysia

²International Institute for Halal Research and Training, International Islamic University
Kuala Lumpur, Malaysia

*Corresponding author: mira@iium.edu.my

(Received: 29th August 2020; Accepted: 5th November 2020;
Published on-line: 30th November 2020)

ABSTRACT: Advanced analytical practices such as data mining or predictive analytics are concepts that are increasingly vital in the area of large data sets. Voluminous data are collected over the years and it is important to assess the data quality for the value of information. Large amounts of data can contain knowledge in the form of patterns. What knowledge an organization especially the high-paced halal industry can get once data quality is assessed and data mining technique is applied? In this research, we have two objectives. Number one is to assess the data quality on the unstructured data collected from the Fourier-transform Infrared Spectroscopy (FTIR) instrument and number two is to identify patterns of halal and non-halal samples which have been analyzed using the FTIR instrument from International Institute for Halal Research & Training (INHART) laboratory. Total Data Quality Management (TDQM) methodology is used in this research where field observation, interview with the stakeholders, and data auditing will be our main objectives. Decision tree for data mining technique is used and this will be done by classifying the sample set based on its halalness.

KEY WORDS: *Data Quality, Data Mining, Halal Industry*

1. INTRODUCTION

The precondition in guaranteeing the best result possible in analyzing data is by having high-quality data (Cai & Zhu, 2015). The result of having poor data quality might turn them to become inaccurate, which may cause on wastage of resources, harming the organization internally, and with its customers (Lucas, 2010). Without solid data, an organization might not reach into a good decision due to lack of understanding on things that are happening within the organization itself (Eckerson, 2002). Eckerson (2002) also describes that several problems that might appear due to poor data quality can include loss in revenue, customer dissatisfaction, loss of credibility in a system, delay in deploying a new system and extra time in data reconciliation. Assessing quality of data is in fact is compulsory to avoid making a wrong decision (Abarza et al., 2014).

Lucas (2010) stated that Data Quality Management (DQM) is a growing issue to the academic and professional communities. The starting point of the furthering data quality is the Data Quality Assessment (DQA) according to Liu et al., (2017). It is crucial to assess data quality before any other related activities which involves

using those data. Ruan et al., (2009) discovered that using regular expressions to analyse and verify the data can help the organization to improve the data quality. The implementation of this, proves the practicality for data quality analysis. Placing a stakeholder in the loop in quality assurance process is another concept for assuring data quality. In which most of the time, users contain the domain-specific knowledge of this application (El Bekri, N. 2016).

Furthermore, ensuring a high data quality level is another vital matter to increase the productivity of the business developments. Having poor data quality might turn these data into incorrect information - which can lead to negative effects for growth of the business. So, by recognizing data quality issues to begin with, this would create a positive effect on efficiency of the business processes (Panahy et al., 2014).

One of the most important concepts in Islam is the concept of Halal, which Halal industry plays a massive significant role in the world, especially to the Muslim society (Abd Aziz et al., 2015). With the continuous increase of global Muslim population, it creates a higher demand for halal food, products, and services (Sulaiman, 2016). As we talk about halal food, the topic of halal market must be taken into the top list of discussion, whereby its logistics and supply chain industry are considered as very dynamic. Consumers demand high quality halal products across the world, and it must have a high attention (Sarah et al., 2011).

International Institute for Halal Research & Training (INHART) is an institution govern under International Islamic University Malaysia (IIUM), which their objective is to assist the government's aspiration to put Malaysia as the global halal hub and to assist halal industry efforts globally.

One of the services which INHART team does is to analyze materials coming from halal and non-halal's samples and transform it to a spectra-based data using Fourier-transform Infrared Spectroscopy (FTIR) instrument. Multiple analysis have been done for the unstructured data but the quality remains unclear.

2. LITERATURE REVIEW

4.1 Halal Industry

Information Technology (IT) has progressed rapidly over the years. Both profit and non-profit organizations (consisting of governmental and non-governmental organizations) began to shift towards the advancement of IT (Tongja et al., 2005). As the Islamic economy continues to grow, it was estimated that Muslim spend US\$1.3 trillion in 2017 alone for the consumption of food and beverages. It was also recorded that Muslim worldwide spend US\$87 billion and US\$62 billion for pharmaceuticals and cosmetics respectively (Reuters & Standard, 2018).

Omar & Jaafar (2011) stated the market of halal products is broadly distributed across the world. This leads to an increase to the demand for halal products – which are not limited to Muslim countries but also from non-Muslim countries. Consequently, this triggers the products to become well-known globally as it highlights the needs for a halal supply chain. Due to this reason, halal supply chain plays a vital part as they underline the concept of halal is not only applicable to the food, but also including the entire supply chain starting from farm sourcing until it reaches to the hands of customers.

Subsequently, several factors have been identified that led to an applicable halal supply chain in relation to the food consumption in Malaysia. Each segment in the halal supply chain is essential on classifying the critical halal control point which can make the halal product turn out to be a non-halal products because of the contamination with the non-halal product on any point of time during storing, packaging, handling and transporting the products (Omar & Jaafar 2011).

4.2 Data Quality Assessment

Data Quality Assessment is being used to present the data whether the quality is good enough to assist people to understand reliability, trustworthiness, and integrity of data quality (Chen et al., 2012). As of today, the data quality issue is increasingly high with unnecessary and dirty data gathering around in information systems (Ruan et al., 2009). According to Kläs et al., (2016), the challenges for data quality assessment can be divided into 3 parts, which are volume, variety, and velocity.

Big data quality evaluation is part of data quality assessment. Serhani et al., (2017) describes that to reach its potentials, researchers must focus on three different areas in big data - data processing, data storage, and data analysis. Hybrid big data quality assessment models are introduced by combining both data-driven and process-driven quality evaluation.

Though data quality assessment has been thoroughly used in many sectors such as businesses, telecommunications, engineering, physics, and healthcare; however, not much research focuses on the halal science. Panahy et al., (2014) suggests that by recognizing the connection between data quality and business processes helps organizations to produce a better decision in the business sector.

Another research for data quality assessment in telecommunication companies is done by Lubis (2012), who describes that quality data maintenance can help to identify customer behaviour and tackle various problems, such as over budget and over capacity within the organization. Liu (2017) stated the implementation of data quality framework in China power grid improved success and dependability of the applications. This was done by focusing on 3 major aspects which are data acquisition, data storage, and data computation.

Data quality plays a major role financially to businesses. Hence, poor data quality can cost the organization between 8% to 12% of its revenue. The results of poor data quality in businesses can have long-lasting harmful effects on the organization's profitability, development and pace at which they operate (Redman, 1998). In describing data quality cost, Eppler et al., (2002) asserts that in terms of data quality, a narrow definition of cost is rather damaging than productive as it carries a negative monetary effect which is caused by inadequate data quality.

Assessing data to observe its quality provides critical context to decision makers and information analysts who look to derive business value from the data. A basic example would be the difference in time that a modelling analysis could be completed with different levels of data quality, where poor quality equates to six weeks and high quality could be six hours (Burkhardt et al., 2018).

Total Data Quality Management (TDQM) is the most used framework to assess the data quality in the organization. Wang (1998) developed the ideas, values, and procedures for calculating, analysing, and improving information

products using Total Data Quality Management (TDQM). Organisation must have a clear articulated Information Products (IP) in business relations, secondly, to establish an IP team consisting of senior executives as the TDQM champion, an IP engineer who is familiar with TDQM Methodology, and members who are information suppliers, manufacturers, etc. Thirdly, to teach Information Quality (IQ) Assessment and IQ Management skills to all the IP constituencies and lastly institutionalize continuous IP Improvement

4.3 Decision Tree

Decision tree is a structure that contains nodes (rectangular boxes) and edges (arrows) and is built from a dataset (table of columns representing features/attributes and rows corresponds to records). Each node is either used to make a decision (known as decision node) or represent an outcome (known as leaf node).

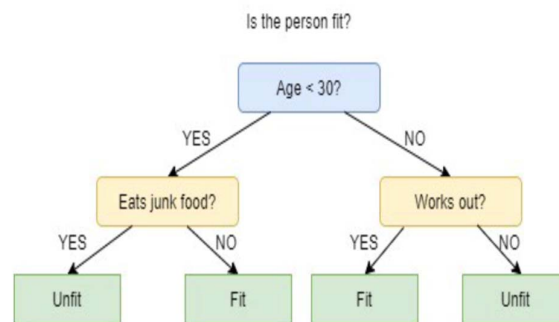


Fig. 1. Decision Tree Example

Fig. 1. explain decision tree used to classify whether a person is fit or unfit. The question nodes will have two different outcomes which is fit and unfit. By looking at the decision tree the following decision can be made: If a person is less than 30 years old and don't eat junk food and that person is fit and if a person is less than 30 years old and eat junk food, then the person is unfit .

ID3 is the most popular algorithm to build a decision tree from a fixed set of sample data. ID3 stands for Iterative Dichotomiser 3 and is named such because the algorithm iteratively (repeatedly) dichotomizes (divides) features into two or more groups at each step. Invented by Ross Quinlan, ID3 uses a top-down greedy approach to build a decision tree. In simple words, the top-down approach means that we start building the tree from the top and the greedy approach means that at each iteration we select the best feature at the present moment to create a node (cise.ufl.edu, 2020)

An obvious advantage of decision trees over other classification approaches is that the tree-like structure allows simplicity of data analysis and interpretation. Decision trees and other classification algorithms by extension pursue to estimate the value of the output attribute based on the values of input attributes. In decision trees, each interior tree node associates to an input attribute. For a nominal interior node, its number of edges is equivalent to the number of feasible values of the related input attribute. Every leaf node in a tree is regarded as a value of the

output attribute inferred from the input attributes and this is represented by the path connecting the root and the leaf (Nwulu 2017).

Some disadvantages of using ID3 algorithm are data may be over-fitted or over-classified, if a small sample is tested as well as only one attribute at a time is tested for making a decision (docs.rapidminer.com 2018).

3. METHODOLOGY

4.1 Total Data Quality Management (TDQM)

Data quality can be translated as a quality of data where the data meets the projected requirements of its use. It is becoming a concern for every organization as the world is moving ahead in the big data era (Abbasi et al, 2016).

In this research, interview methods are used as well as observations methods. FTIR data results data were collected from INHART's lab. 223 unique objects samples were collected from 419 sets of data. Multiple file's format have been received during the interview from three different researchers such as CSV and PDF file.

During the observation methods, INHART's researchers performed how to analyze an object and transform them into a graphical spectrum. A sample object needs to be run at least five times for the accuracy purposes. In some occasions, a sample could be processed 60 times to make sure the accuracy is higher.

The FTIR instrument could provide the result of the spectrum in many different file formats. INHART's researchers in particular are always transforming the results file into Comma Separated Value (CSV) file and Portable Document Format (PDF) file. Spectrum files from Absorbance and Wavenumbers are translated into a number for a given range and it consists of 5,000 rows of data. Additionally, PDF file format gives a graphical look on the spectrum of the data.

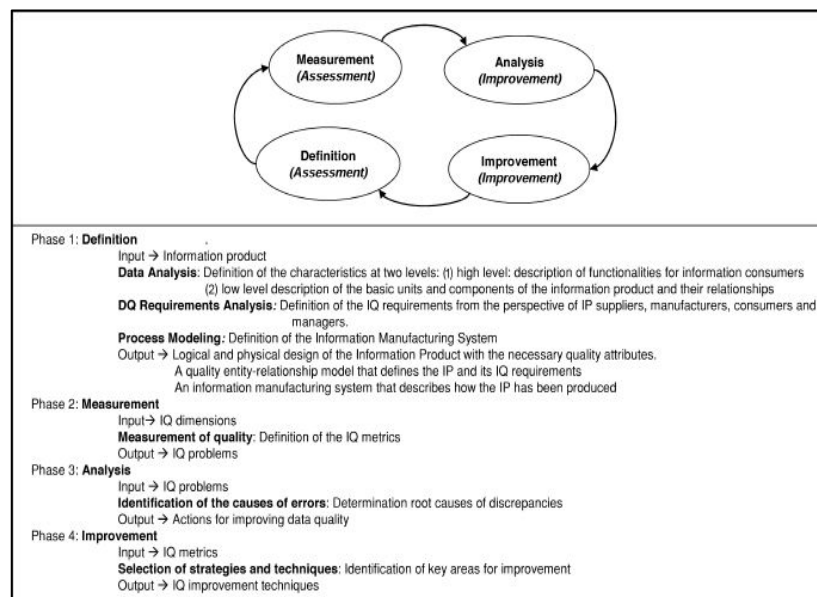


Fig. 2. TDQM Methodology (Batini et al., 2009)

TDQM has been used to eliminate discrepancies between the output of the device processes and its requirements from the users – in which its methodology is shown on the figure 1.3 above (Batini et al., 2009). TDQM carries the goal of providing support to the quality improvement development from requirements to implementation. Several data quality dimensions are also tackled in this method such as, accessibility, appropriateness, completeness, relevance, reputation, timeliness, and understandability.

4.2 Cross Industry Standard Process For Data Mining (CRISP-DM)

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a consortium of companies established by the European Commission in 1996 and has been designated as standard processes in data mining that can be applied at various industrial sectors.

This methodology is also broadly used to build mining models especially for data mining predictive modelling methods. It gained its acceptance among the other standards (SEMMA and PMML) in predictive data mining which explain the process of data mining. Shearer et al. (2000) defined CRISP-DM as a non-proprietary, documented, and freely accessible data mining model which is beneficial for the organizations since this methodology offers best practices, better, and faster results for data mining (Olson & Delen., 2008).

The CRISP-DM breaks data mining process into 6 phases: business and data understanding, data preparation, modelling, evaluation, and deployment. These phases assist organizations to comprehend the data mining process and deliver a road map as guidance while planning and performing a data mining task. Below is a figure of the life cycle development process of CRISP-DM (Olson & Delen., 2008).

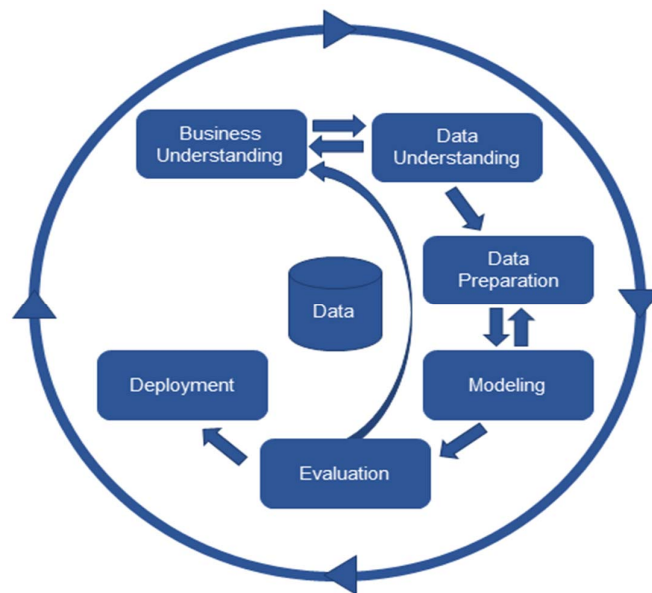


Fig. 3. CRISP-DM phases in data mining (Olson & Delen., 2008)

The CRIPS-DM consists of six phases (Olson & Delen., 2008).

1. **Business Understanding.** It contains deciding business objectives, assessing the present circumstance, building up data mining objectives, and building up a task plan.
2. **Data Understanding.** Once business objectives are recognized, we continue with data requirements. During this phase we will do data collection, data description, data exploration, and the verification of data quality.
3. **Data Preparation.** This phase will be focusing on data selection, data cleaning, as well as data exploration. This phase will provide the opportunity to see patterns based on business understanding.
4. **Modeling.** Data mining tools like RapidMiner are in place for modeling. During this phase, greater data understanding is gained, and more detailed appropriate models can be applied.
5. **Evaluation.** In this phase, model results will be evaluated in context of the business objectives. This will prompt the recognizable pieces of proof of different needs.
6. **Deployment.** The results of the data mining study are concluded, and it needs to be reported back to the stakeholders. It will reveal new information which needs to be tied to the original data mining project goals.

4. RESULTS

4.1 Data Quality Assessment Results

Three Information Quality (IQ) metrics is defined for the result which includes: availability, usability, and, reliability.

4.1.1 Availability

The first dimensions in the hierarchical big data quality assessment framework is Availability. It has two dimensions which are accessibility and timeliness. In INHART itself, the data repository for FTIR results is not presented in a form where every user can access it. The lack of an online database to access every data is not implemented yet. Therefore, data is collected from a known individual researcher only which was recommended by INHART's lab assistant.

The process of running the sample in an FTIR instrument is time consuming due to its preparation steps. When the sample data from FTIR has shown its results, storing it directly in the database in real-time can be fundamental. To precisely record every results data to the data repository could credibly find the optimum collaboration between INHART's researchers. This would tackle any duplication in work for their future reference.

In a dynamic and rapidly evolving system, a second or even a few microseconds between one reading and another can mean one dataset is mismatched against another. It's essential to ensure that all data is timely and being interpreted in real-time to ensure the best optimization of the system.

4.1.2 Usability

Usability dimension has the credibility elements attached to it. Data credibility is the level of integrity in a data provider or source that can be relied upon, to ensure that the data really represents what the data is supposed to represent, and there is no intent to misrepresent what the data is supposed to represent.

Data integrity is critical for combined activities where organizations share information within and across the organizations so that decision makers can analyze and mine the data. The difficulty of data integrity becomes predominantly challenging for large-scale collaborations in which parties make frequent data modifications. Without integrity, collaborations among multiple user in organization would not be successful since the usefulness of data becomes depreciated and it cannot be trusted with confidence (Lim et al., 2010)

The data that the INHART's teams produced are reliable due its worldwide recognition for their service in halal industry. We noticed also that only researchers or students from INHART that could only access and run the data analysis for FTIR. This gives credibility for INHART's brand because only the people who are associated with INHART that can run the experiment.

4.1.3 Reliability

Data accuracy, consistency, integrity, and completeness are the elements for data reliability dimensions. At the moment, the end process to get the FTIR result data is not inclusive yet. This happens because researchers need to run several experiments in order to the high level of accuracy of FTIR spectra's result. There is no assessment being done yet for accuracy of data because the results actually depend on how the sample is being prepared. The pre-processing of sample objects fully depends on how cautious the researchers are in following the Standard Operational Procedure (SOP).

The data sources are represented as a collection of data sets or files, diversified or arranged by types, format or domain. This representation of data will assist to avoid quality issues that may appear when auto-discovery of data types and domain is handled.

4.2 Decision Tree

Below is the sample data taken from FTIR,

	A	B	C	D	E	F	G
1	Obserbance	Wavenumbers	Object				
2	0.1655795	649.904	PRCN1				
3	0.1654403	650.3861	PRCN1				
4	0.1655249	650.8682	PRCN1				
5	0.1658798	651.3503	PRCN1				
6	0.1664004	651.8325	PRCN1				
7	0.1668802	652.3146	PRCN1				
8	0.1671181	652.7968	PRCN1				
9	0.1670242	653.2789	PRCN1				
10	0.1666617	653.761	PRCN1				
11	0.1661976	654.2431	PRCN1				
12	0.1658024	654.7252	PRCN1				
13	0.1655556	655.2073	PRCN1				
14	0.1654109	655.6895	PRCN1				
15	0.1652387	656.1716	PRCN1				
16	0.1649083	656.6537	PRCN1				
17	0.1643777	657.1359	PRCN1				
18	0.1637221	657.618	PRCN1				
19	0.1630945	658.1001	PRCN1				
20	0.1626408	658.5822	PRCN1				
21	0.1624224	659.0643	PRCN1				

Fig. 4. Example of FTIR results in Comma Separated Values (CSV) format.

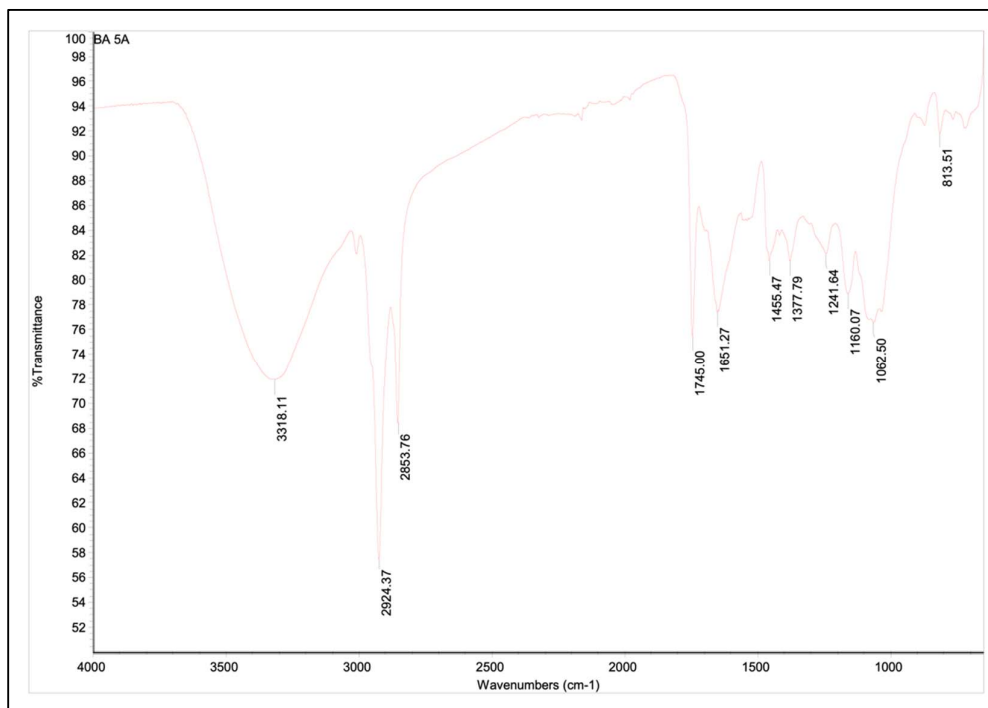


Fig. 5. Example of FTIR results in graph format

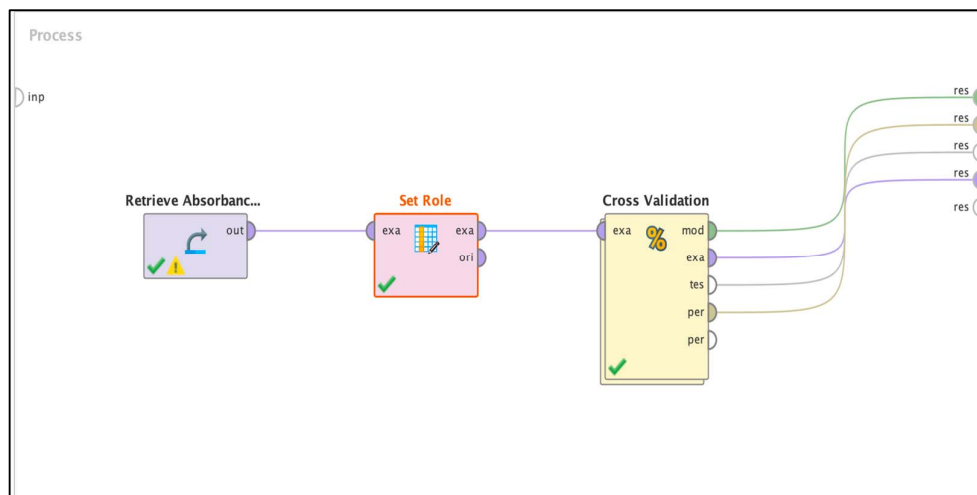
Table 1: Example Set of data mining processing of Decision Tree FTIR Result for RapidMiner Studio version 9.7.000

Row No	Sample Name	Absorbance	Halalness
2	PRCN	0.3487459	No
6952	FSH	0.3929258	Yes
13902	GOAT	0.2458752	Yes
28113	PO 1	10.5	No
20852	PO 14	17.4	No
42324	BVN	0.1457702	Yes

Decision tree is chosen for data mining activity due to its approaches which allows simplicity of data analysis and interpretation. RapidMiner Studio version 9.7.000 is also used as a tool for data mining.

Table 1 will be the baseline for the process of decision tree. It shows the decision system which will be processed in this study. There are many samples which falls under halal category during the data collection. We've selected three samples (FSH, GOAT, and BVN) which are the animal sample which structure are similar to pork.

Below is the process on how the we run the analysis of decision tree using RapidMiner:



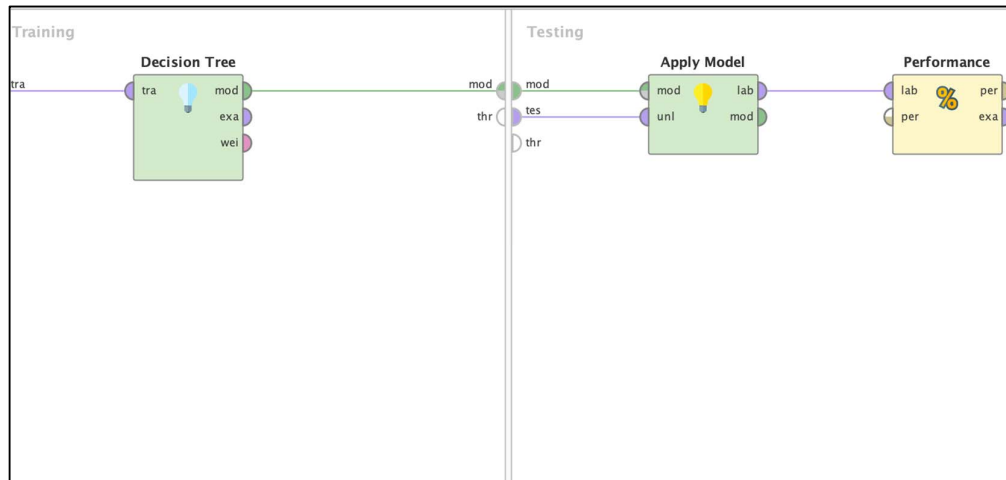


Fig. 6. Model processing for Decision Tree operator and Cross Validation operator in RapidMiner 9.7.000

The first entry point is by having the dataset with multiple samples from the halal and the non-halal samples. The sample set that we put in this analysis are coming from three halal sample from animal based such bovine, goat, and fish. While the other three samples from the non halal animal based are porcine, combination of *palm oil + pork + chicken*, and combination of *palm oil + onions + pork*. The total number of 49,000 rows of data are used on this process

Secondly, we add another operator named *Set Role* which is used to change the role of one or more attributes. The role of an attribute describes how other operators handle this attribute. The default role is regular, other roles are classified as special. An ExampleSet can have many special attributes, but each special role can only appear once (RapidMiner 2020). We target the role as label and attribute it to sample's name. We train the model to predict the sample's name so it is set to the label role.

Thirdly, another operator called *Cross Validation* which is used to perform a cross validation to estimate the statistical performance of a learning model. We used this to estimate how accurately a model perform in practice. The evaluation of the performance of a model on independent test sets yields a good estimation of the performance on unseen data sets. It also shows if 'overfitting' occurs. This means that the model represents the testing data very well, but it does not generalize well for new data. Thus, the performance can be much worse on test data.

We also put *Decision Tree* operator on the training model inside the *Cross Validation* operator which will generate a decision tree model based on the samples that we put on the data.

Based on the sample set data statistics, the training data is actually close to perfect because there is no missing value for all the samples. In a lot of cases algorithm cannot handle the missing values but in this case the our data set is

close to perfect and the accuracy of the decision tree which we will perform would be higher. We set the Name column as a label and the absorbance and halalness is assigned to attributes.

accuracy: 64.16% +/- 1.04% (micro average: 64.16%)

	true PRCN	true FSH	true GOAT	true PO 14	true PO 1	true BVN	class precision
pred. PRCN	13900	0	0	3	7	0	99.93%
pred. FSH	0	2662	645	0	0	68	78.87%
pred. GOAT	0	2468	3725	0	0	2822	41.32%
pred. PO 14	0	0	0	7251	7240	0	50.04%
pred. PO 1	0	0	0	7	14	0	66.67%
pred. BVN	0	1820	2580	0	0	4060	47.99%
class recall	100.00%	38.30%	53.60%	99.86%	0.19%	58.42%	

Fig. 7. Performance Vector results

In this performance vector analysis we have found out that the decision tree model that we run have accuracy of 64 %. But the most interesting part is for Porcine have 99% accuracy compared to the other samples.

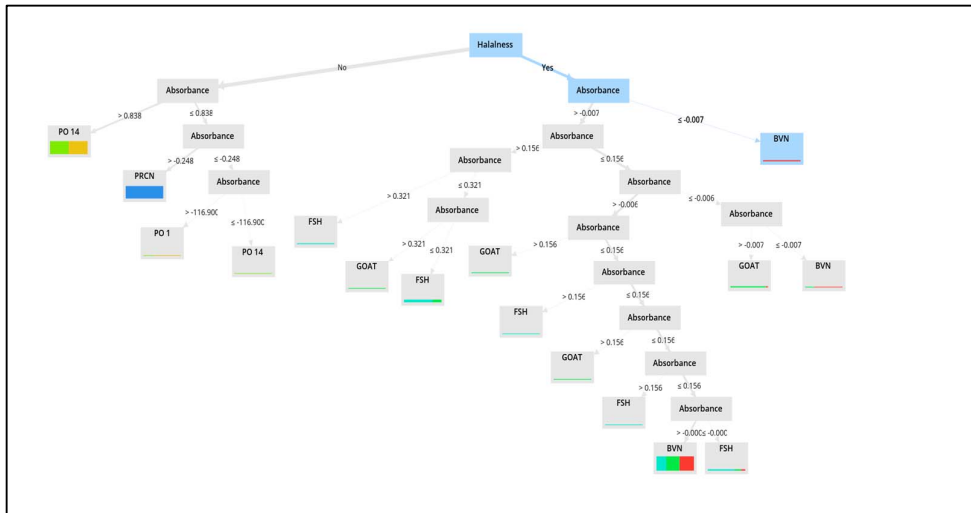


Fig. 8. Decision Tree result of the sample set

The result for the decision tree model here tells that if the sample of the absorbance is less than 0.838 it is classified under non-halal result. The thickness of the coloured bar indicates out of how many example the branch has made.

As we can see if the halalness is equals to No then we can see the absorbance will be greater than 0.838. This mainly coming from Porcine sample set. Another interesting fact here is that if the absorbance is less than or equals to 0.838 it is fall under non-halal sample such PO14 and PO1.

We also can see that if the spectrum is greater than -0.007 it will falls into halal samples such Goat, Fish, and Bovine

5. CONCLUSION AND FUTURE WORKS

INHART's team might be able to find a similar replacement of non-halal gelatin which currently are used widely for cosmetic products. The result of decision tree shows that the current sample which coming from INHART's lab got no similarities in its spectra number. Also, the team might be able to categorize samples coming from FTIR instrument according to its own nature. In conclusion, when the factors influencing the data mining performance are observed, data mining can contribute more in halal industry by providing patterns.

Furthermore, INHART's team might want to focus on examining more of the non-halal sample due to the limitations of the data collected from the lab. By doing this, we can make more analysis that is not limited to 3 data sample only and the clarity of the result should be benefited INHART's team for further analysis.

This study faced several challenges and limitations during its research. For further research, we would like to recommend to use another method of data mining technique. Only one data mining technique were used in this perform data mining in this study, it would be better in the future research if we can compared between two or three techniques so it can be identified which technique could generate most accurate prediction and the most suitable technique can be tested on FTIR spectra results.

ACKNOWLEDGEMENT

The authors would like to thank KIHIM for funding this research (Grant No. MOHE 18-002-0002) and Dr Mohamed Elwathig Mirghani of INHART for assisting in FTIR data interpretation.

REFERENCES

- Abarza, R. G., Motz, R., & Urrutia, A. (2014). Quality Assessment Using Data Ontologies. *2014 33rd International Conference of the Chilean Computer Science Society*.
- Abarza, R. G., Motz, R., & Urrutia, A. (2014). Quality Assessment Using Data Ontologies. *2014 33rd International Conference of the Chilean Computer Science Society*.
- Abbasi, A., Sarker, S. and Chiang, R. H., (2016). Big Data Research in Information Systems.
- Abbasi, A., Sarker, S. and Chiang, R. H., (2016). Big Data Research in Information Systems.
- Abd Aziz, N. N., Abd Aziz, N. A., Abd Aziz, N. A., Omar, Z., & Aziah Wan Hassan, W. H. (2015). A Review on the Emergence and Growth of Halal Studies. INTERNATIONAL ACCOUNTING AND BUSINESS CONFERENCE, IABC 2015.
- Abd Aziz, N. N., Abd Aziz, N. A., Abd Aziz, N. A., Omar, Z., & Aziah Wan Hassan, W. H. (2015). A Review on the Emergence and Growth of Halal Studies. INTERNATIONAL ACCOUNTING AND BUSINESS CONFERENCE, IABC 2015
- Batini, C., Francalanci, C., Cappiello, C., & Maurino, A. (2009). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*.

- Batini, C., Francalanci, C., Cappiello, C., & Maurino, A. (2009). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*
- Burkhardt, A., Berryman, S., Brio, A., Ferkau, S., Hubner, G., Lynch, K., Sonderer, K. (2018). Measuring Manufacturing Test Data Analysis Quality.
- Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14: 2, Pp. 1-10.
- Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14: 2, Pp. 1-10.
- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health*.
- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health*
- Eckerson, W. W. (2002). Data Quality and The Bottom Line. *TDWI Report Series*.
- Eckerson, W. W. (2002). Data Quality and The Bottom Line. *TDWI Report Series*
- El Bekri, N., & Peinsipp-Byma, E. (2016). Assuring Data Quality by Placing the User in the Loop. *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, 468-471.
- El Bekri, N., & Peinsipp-Byma, E. (2016). Assuring Data Quality by Placing the User in the Loop. *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, 468-471.
- Eppler, M., & Muenzenmayer, P. (2002). Measuring Information Quality in The Web Context: A Survey of State-Of-The-Art Instruments and An Application Methodology. *Proceedings of the Seventh International Conference on Information Quality*.
- Kläs, M., Putz, W., & Lutz, T. (2016). Quality Evaluation for Big Data. *Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement*.
- Kläs, M., Putz, W., & Lutz, T. (2016). Quality Evaluation for Big Data. *Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement*.
- Lim, H. S., Dai, C., & Bertino, E. (2010). A Policy-based Approach for Assuring Data Integrity in DBMSs. *6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2010) Chicago*.
- Liu, H., Huang, F., Li, H., Liu, W., & Wang, T. (2017). A Big Data Framework for Electric Power Data Quality Assessment. *2017 14th Web Information Systems and Applications Conference. (WISA)*, 289-292.
- Liu, H., Huang, F., Li, H., Liu, W., & Wang, T. (2017). A Big Data Framework for Electric Power Data Quality Assessment. *2017 14th Web Information Systems and Applications Conference. (WISA)*, 289-292.
- Lubis, M., & Kartiwi, M. (2012). Data Quality Assessment on Internal System in XYZ Division of Telecommunication Company: A Case Study in Indonesia.
- Lucas, A. (2010). Corporate Data Quality Management from Theory to Practice. *5th Iberian Conference on Information Systems and Technologies*, 1-7.
- Lucas, A. (2010). Corporate Data Quality Management from Theory to Practice. *5th Iberian Conference on Information Systems and Technologies*, 1-7.
- Mohd Sulaiman, M. Z., Noordin, N., Md Noor, N., Hakim Suhaimi, A. I., & Wan Mohd Isa, W. (2016). Case of Halal Certification in Malaysia: A Case of Halal Certification in

- Malaysia. *2016 6th International Conference on Information and Communication Technology for The Muslim World*.
- Nwulu, N. I. (2017). *A Decision Trees Approach to Oil Price Prediction*. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. doi:10.1109/IDAP.2017.8090313
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques (XII ed., Vol. 1, 180 p)*. ISBN:978-3-540-76916-3
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques (XII ed., Vol. 1, 180 p)*. ISBN:978-3-540-76916-3
- Omar, E. N., & Jaafar, H. S. (2011). Halal supply chain in the food industry - A conceptual model. *2011 IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA), Langkawi, Malaysia*. doi:10.1109/ISBEIA.2011.6088842
- Omar, E. N., & Jaafar, H. S. (2011). Halal supply chain in the food industry - A conceptual model. *2011 IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA), Langkawi, Malaysia*. doi:10.1109/ISBEIA.2011.6088842
- Panahy, P. H., Sidi, F., Affendey, L. S., & Jabar, M. A. (2014). The Impact of Data Quality Dimensions on Business Process Improvement. *2014 4th World Congress on Information and Communication Technologies (WICT 2014)*, 70-73.
- Panahy, P. H., Sidi, F., Affendey, L. S., & Jabar, M. A. (2014). The Impact of Data Quality Dimensions on Business Process Improvement. *2014 4th World Congress on Information and Communication Technologies (WICT 2014)*, 70-73.
- Patil, D. V., & Bichkar, R. S. (2006). A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets.
- Patil, D. V., & Bichkar, R. S. (2006). A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets.
- Patil, S., & Kulkarni, U. (2019). Accuracy Prediction for Distributed Decision Tree using Machine Learning approach. *Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)*.
- Rangra, K., & Bansal, K. L. (2014). Comparative Study of Data Mining Tools. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(6).
- Rangra, K., & Bansal, K. L. (2014). Comparative Study of Data Mining Tools. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(6).
- RapidMiner Studio System Requirements. (n.d.). Retrieved December 22, 2018, from <https://docs.rapidminer.com/latest/studio/installation/system-requirements.html>
- RapidMiner Studio System Requirements. (n.d.). Retrieved December 22, 2018, from <https://docs.rapidminer.com/latest/studio/installation/system-requirements.html>
- Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), 79-82.
- Reuters, T., & Standard, D. (2018). State of the Global Islamic Economy Report 2018/19.
- Reuters, T., & Standard, D. (2018). State of the Global Islamic Economy Report 2018/19.
- Ruan, H., Yu, D., & Cao, Y. (2009). Research and Implementation of the Platform for Analyzing Data Quality. *2009 Second International Workshop on Computer Science and Engineering*, 1, 274-278.
- Ruan, H., Yu, D., & Cao, Y. (2009). Research and Implementation of the Platform for Analyzing Data Quality. *2009 Second International Workshop on Computer Science and Engineering*, 1, 274-278.

- Sarah, S., Illyas, M. I., & Desa, M. I. (2011). Tracking and Tracing Technology for Halal Product Integrity over the Supply Chain. *2011 International Conference on Electrical Engineering and Informatics*, 1-7.
- Sarah, S., Illyas, M. I., & Desa, M. I. (2011). Tracking and Tracing Technology for Halal Product Integrity over the Supply Chain. *2011 International Conference on Electrical Engineering and Informatics*, 1-7.
- Serhani, M. A., El Kassabi, H. T., Taleb, I., & Nujum, A. (2016). An Hybrid Approach to Quality Evaluation Across Big Data Value Chain. *2016 IEEE International Congress on Big Data*.
- Serhani, M. A., El Kassabi, H. T., Taleb, I., & Nujum, A. (2016). An Hybrid Approach to Quality Evaluation Across Big Data Value Chain. *2016 IEEE International Congress on Big Data*.
- The ID3 Algoritihm. Retrived November 16, 2020, from <https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm>
- Tongja, R., Subrahmanian, E., Arunachalam V.S. (2005). Information and Communication Technology for Sustainable Development: Defining a Global Research Agenda.
- Tongja, R., Subrahmanian, E., Arunachalam V.S. (2005). Information and Communication Technology for Sustainable Development: Defining a Global Research Agenda.
- Wang, R. (1998). Total Data Quality Management.
- Wang, R. (1998). Total Data Quality Management.
- Yang, F. J. (2019). An Extended Idea about Decision Trees. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*.
- Yang, F. J. (2019). An Extended Idea about Decision Trees. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*.