

Understanding of Data Mining in Computer Science

Learning from PILKADA DKI Jakarta 2017

Harco Leslie Hendric Spits Warnars¹, Ford Lumban Gaol², Agung Trisetyarso³, Edi Abdurachman⁴
 Doctor of Computer Science
 Bina Nusantara University
 Jakarta, Indonesia
 shendric@binus.edu¹, fgaol@binus.edu², atrisetyarso@binus.edu³,
 ediA@binus.edu⁴

Dedy Prasetya
 Kristiadi
 Computer Systems
 STMIK Raharja
 Tangerang, Indonesia
 dedy.prasetya@raharja.info

Tokuro Matsuo
 Graduate School of Industrial Technology
 Advanced Institute of Industrial
 Technology
 Tokyo, Japan
 matsuo@tokuro.net

Abstract—DKI Jakarta gubernatorial election 2017 was phenomenal event which draws attention worldwide, where one of the defeated candidate Ahok or recognized as Basuki Tjahaja Purnama was charged with blasphemy, has double minority background in Indonesia as descendants of China and Christian and he was titled “Man of the Year” 2015 by Globe Asia magazine. However, in this paper we do not interested with political conversation, how one candidate couples can win or how one candidate couples could lose. We are interested to learn based on our science knowledge in Computer Science perspective particularly with the term of Data Mining with this 5 years DKI Jakarta event program. Hopefully, by reading this paper the students of Computer Science can understand their Computer Science knowledge in their surrounding life events. Obviously, voters’ decisions are influenced by human nature characteristic which like to similarity preferences, where human like to choose someone who have full or frequent similarity in profile and/or program. Beforehand, voters should do learning by doing the characterization of their and candidate couples’ profile and program. Afterwards, based on candidate couples’ profile and program, the voters will do discrimination between candidate couples, classification, clustering or association among candidate couples. Admittedly, the voters will fill their ballot based on their preference similarity.

Keywords—Data Mining; Artificial Intelligent; characteristic rule; discriminant rule; classification rule;

I. INTRODUCTION

Pilkada (Pemilihan umum Kepala dan wakil kepala Daerah) DKI (Daerah Khusus Ibukota) Jakarta 2017 is phenomenal event in Indonesia or even worldwide. This DKI Jakarta governor election were running in 2 rounds, where first round which run on 15 February 2017 had 3 candidate couples such as Agus-Silvi (Agus Harimurti yudhoyono, M.Sc, M.P.A, M.A. and Prof. Dr. Hj. Sylviana Murni, S.H., M.Si), Ahok-Jarot (Ir.Basuki Tjahaja Purnama, M.M. and Drs.H. Djarot Saiful Hidayat,M.S.) and Anies-Sandiaga (Anies Rasyid Baswedan, Ph.D and Sandiaga Salahuddin Uno, B.A., M.B.A) with percentage count result 17.06%, 42.96% and 39.97% respectively[5]. The second round which is run on 19 April 2017 had 2 candidate couples such as Ahok-Jarot and Anies-Sandiaga with percentage count result 42.05% and 57.95% respectively[6]. In first round of Jakarta gubernatorial election 2017, Agus-silvi was supported by 4 parties such as Demokrat, PPP, PKB and PAN where each of them have number of seats

in regional parlement (DPRD=Dewan Perwakilan Rakyat Daerah) such as 10, 10, 6 and 2 respectively with total 28 seats. Meanwhile, Ahok-Jarot was supported by 4 parties such as PDIP, Hanura, Golkar and NasDem with number of seats such as 28,10, 9 and 5 respectively with total 52 seats. Futhermore, Anies-Sandiaga was supported by only 2 parties such as GERINDRA and PKS with number of seats 15 and 11 respectively with total 26 seats. Figure 1 shows the 1st round of result and composition of DKI Jakarta gubernatorial election.

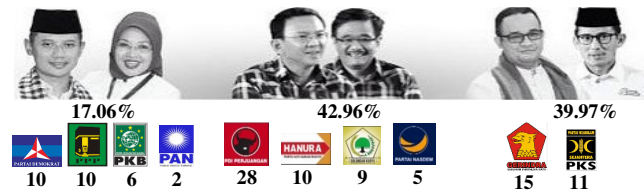


Fig. 1. 1st round of DKI Jakarta gubernatorial election result and composition

Furthermore, in second round, candidate couple Ahok-Jarot was get additional supported from 2 parties such as PPP and PKB with total 72 seats of regional parlement, where PAN party supports to candidate couple Anies-Sandi with total 32 seats of DPRD, and Demokrat becomes neutral party. Figure 2 shows the 2nd round of result and composition of DKI Jakarta gubernatorial election.



Fig. 2. 2nd round of DKI Jakarta gubernatorial election result and composition

Meanwhile, in Computer science, there is technology which called Data Mining or have other name recognition such as Knowledge Data Discovery (KDD), Machine Learning, Deep Learning, statistics, pattern recognition and so on. Data Mining is computer algorithm program which is part of Artificial Intelligent (AI) where have meaning to translate human thinking and acting. AI that translates human thinking into computer programming algorithms has embodied in technologies such as Data Mining, KDD, Machine Learning, Deep Learning, Statistics, Decision Support System (DSS), Dashboard, Business Intelligent (BI), Knowledge Management

(KM) and so on. This AI human thinking computer algorithm program is run in gadget such as smart phone/ laptop/computer/tablet and so on. This AI technologies can be applied in either Web or non web based application with language programming such as HTML, JSP, PHP, ASP or C,C++,C#, Java, Python, .NET respectively.

Meanwhile, AI that translates the way humans act into the computer program algorithm has embodied in robot research, including your printer. This AI human thinking computer algorithm program is embedded/injected in microcontroller which is mounted on Printed Circuit Board (PCB), include motor controller and some optional attached sensors such as light sensor, temperature sensor, smoke sensor, Gas sensor, pressure sensor, touch sensor, Infrared sensor, ultrasonic sensor, speed sensor, humidity sensor, motion sensor, distance sensor and so on. The robot can be controlled with either wired or wireless connection. Microcontroller can be programmed with programming software such as assembler, C, C++, C#, Python, Java, .NET, Arduino software and so on.

Thus, since Data Mining is the translation of human thinking into the computer program algorithm, so inevitably the process of governor elections DKI 2017 which involves many heads, will automatically involve many ways of human thinking. Obviously, the human way of thinking in this Pilkada DKI 2017 process can serve as an example for understanding the theory of Data Mining in computer science.

II. SIMPLE DEFINITION ABOUT DATA MINING THEORY

Before we continue, then better to understand what is Data Mining all about. Data Mining is computer algorithm program which have function to find or learn pattern/rule in data as a explicit knowledge in Knowledge Management. There are many current recognized Data Mining algorithms such as Decision tree/rules, Support Vector Machine (SVM), C4.5, K-Means, Apriori, Expectation Maximization (EM), PageRank, AdaBoost, k-Nearest Neighbors (kNN), Naive Bayes and Classification and Regression Tress(CART)[2,3], Emerging Pattern (EP)[7-8], Attribute Oriented Induction (AOI) [9-16], Attribute Oriented Induction High Emerging Pattern (AOI-HEP)[17-21]. The pattern or rule is the way to look at the data in many perspective. In Data Mining, there are 5 basic knowledges pattern/rule which can be mined/learned from data as seen in figure 3 and they are:

- a. Characteristic pattern/rule, is identification of 1 only data.
- b. Discriminant pattern/rule, is differentiation between 2 data.
- c. Classification pattern/rule, is differentiation among data.
- d. Association pattern/rule, is implication (→) relationship among data.
- e. Clustering pattern/rule, is relationship among data.

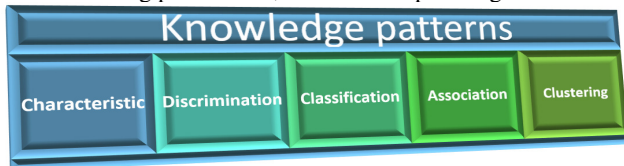


Fig. 3. Human's compartmentalization

It is possible to show quantitative score for all knowledge patterns. Classification and clustering patterns are similar pattern but have different ways to classify the data where classification is supervised whilst clustering is unsupervised. Meanwhile, in Data Mining there are 3 ways to learn knowledge pattern/rule such as characteristic, discriminant, classification, association and clustering and they are[1]:

- a. supervised learning, is known output variable in dataset and input labeled data.
- b. semi supervised learning, whereby the output variable is known only for some examples and input both labeled and unlabeled data.
- c. unsupervised learning, is without known output variable in dataset and input unlabeled data.

However, we just only discuss only supervised and unsupervised learning, where supervised and unsupervised learning with and without human intervention respectively. As lecturer, when we attend to our class and see the way the students get their seated position in class then we have student clustering and it is unsupervised learning, where we have never supervised the students seated position in class. However, when we ask students to fill their seated position in class then we have student classification and it is supervised learning where we supervise the students seated position in class.

III. LEARNING FROM PILKADA DKI JAKARTA 2017

Conscious or unconscious, we are human always do compartmentalizes based on closest similarity between us and other human. We always choose similarity between us and others with similarity such as gender, blood line, race, religion, brotherhood and so on. Figures 4 and 5 as examples of human compartmentalization of 3 candidate couples for 2017 Jakarta gubernatorial election. Figure 4 shows the classification among 3 candidate couples, where there is classification based on closest similarity either full or frequent similarity, between each candidate couple or all candidate couples.

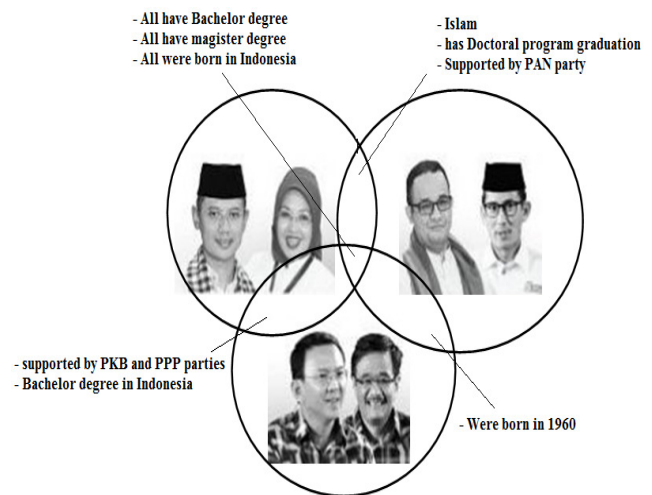


Fig. 4. Classification of Candidate couples of 2017 Jakarta Gubernatorial election

Moreover, figure 4 shows the similarity all candidate couples in having bachelor and master degrees and were born in Indonesia. Moreover, the discriminant between candidate couples which is shown in figure 5 (d) (e) (f) and as shown in figure 4 are similarity between candidate couple Agus-Sylvi and Anies-Sandi in similar Religion=Islam, one of member candidate couple has doctoral degree program and both of them is supported by PAN party. Similarity between candidate couple Anies-Sandi and Ahok-Jarot is only were born in 1960, and similarity between candidate couple Ahok-jarot and Agus-Sylvi which were supported by PKB and PPP parties and having bachelor degree from Indonesia university.

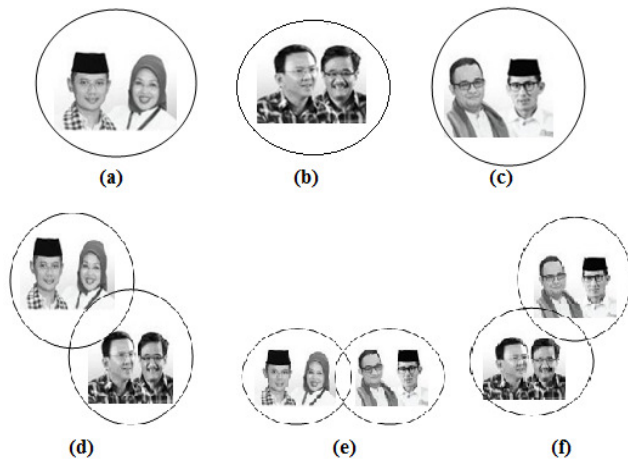


Fig. 5. Characterization and discrimination of Candidate couples of 2017 Jakarta gubernatorial election

Furthermore, figure 5 (a), (b), (c) are examples of characterization of each candidate couples where each candidate can have their own characterization profile as shown in tables 1 and 2[4]. Meanwhile, the voters chose their choice in ballot based on closest similarity on candidate couples' profile and vision-mission. The voters chose their preference candidate couples because their similarity in profile such as gender, religion, race, education and of course they chose because of interest to their programs.

Hereinafter, table 1 and 2 show the profile of candidate couples from website of General Election Commissions Province of DKI Jakarta (<https://kpujakarta.go.id>)[4]. Table 1 shows attributes such as name, R=Religion (1=Christian, 2=Islam), DOB=Date of Birth, POB= Place of Birth, NP= Number of Party, NS=Number of Seats in DPRD, Bachelor Program and Bachelor university. Moreover, table 2 is continuation data from table 1 and has attributes such as name, master program, master university, doctoral program and doctoral university. Meanwhile, in attributes such as bachelor university, master university and doctoral university there are code such as ID, USA and SG where ID=Indonesia, USA=United States of America, SG=Singapore. DOB attribute in table 1 has format data such as the first 2 digit for date, the next 2 digit for monthly digit and the last 4 digit for year. Since there are candidate couples' characterization as seen in tables 1 and 2, then the candidate couples' programs should be

characterized and shown just like tables 1 and 2, in order to extend the process knowledge pattern discovery.

TABLE I. CANDIDATE COUPLES' CHARACTERIZATION PROFILE

Name	R	DOB	POB	NP	NS	Bachelor Program	Bachelor University
Agus	2	1008 1978	Bandung	4	28	Military	Akademi Militer, ID
Sylvi	2	1110 1958	Jakarta	4	28	Law of state administration	Jayabaya univ, ID
Ahok	1	2906 1966	Manggar	6	72	Geology	Trisakti univ, ID
Jarot	2	0607 1962	Mage lang	6	72	Administrati on	Brawijaya univ, ID
Anies	2	0705 1969	Kuni ngan	3	32	Economy	Gajah Mada univ, ID
Sandi	2	2806 1969	Rum bai	3	32	Business Administrati on	Wichita State univ,USA

TABLE II. CONTINUE CANDIDATE COUPLES' CHARACTERIZATION PROFILE

Name	Master Program	Master University	Doctoral Program	Doctoral University
Agus	Strategic Studies	Nanyang Tech univ, SG		
Sylvi	Population Management	Univ Negeri Jakarta, ID	Education managemen t	Univ Negeri Jakarta, ID
Ahok	Management	Prasetya Mulya univ, ID		
Jarot	Politics	Gajah Mada univ, ID		
Anies	Int Security and Eco. Policy	Univ of Maryland, USA	Politics	Nothern Illinois univ, USA
Sandi	Business Administration	George Washington univ, USA		

The implementation of those 5 knowledge patterns/rules such as characteristic, discriminant, classification, association and clustering will split into supervised and unsupervised, where supervised with human intervention whilst another not. Next, are reality life example from Pilkada DKI Jakarta 2017, how those 5 knowledge patterns are applied between supervised and unsupervised in qualitative rule since there is no score number percentage. The presentation pattern/rule will be read based on candidate couples' profile in tables 1 and 2 and as shown in figure 4. Next, are example characteristic pattern/rule for each candidate couples based on their profile in tables 1 and 2 and as shown in figure 4.

- First candidate couple (Agus and Sylvi) have bachelor and master degree, both are Islam, born in Indonesia, born in non 1960, born in Java Island, supported by Demokrat party only, supported by 4 parties, both have bachelor degree in Indonesia, one of them has doctoral degree in Indonesia, one of them has master degree in Singapore (outside Indonesia).
- Second candidate couple (Ahok and Jarot) have bachelor and master degree in Indonesia, has mixed religion Christian and Islam, born in Indonesia, born in 1960, born in Java and Sumatera Islands, supported by 6 parties, none have doctoral degree.
- Third candidate couple (Anies and Sandi) have bachelor and master degree, both are Islam, born in Indonesia, born

in 1960, born in Java and Sumatera Islands, supported by 3 parties, both have bachelor and master degree, one of them has USA (non Indonesia) bachelor degree, both have USA (non Indonesia) graduation degree, both have USA (non Indonesia) master degree, one of them has USA (non Indonesia) doctoral degree.

IV. CONCLUSION

The finding data mining pattern knowledge patterns/rules such as characteristic, discriminant, classification, association and clustering can be defined with formulation as qualitative rule which have score number percentage including growthrate formula as justification the score of how many times and confidentiality as justification of finding pattern/rule. For example as shown for discrimination below, where for discrimination between Ahok and Anies where for higher education(x) profile, Ahok has master degree with score 2, whilst Anies has doctoral degree with score 3. For discrimination Ahok to Anies there is $GR = \text{growthrate} = 0.67$, which means Ahok 0.67 more lower higher education rather than Anies. Meanwhile, discrimination Anies to Ahok there is $GR = \text{growthrate} = 1.5$, which means Anies 1.5 more better in higher education rather than Ahok.

$V(x) = \text{ahok}(x)[2] - \text{Anies}(x)[3] \rightarrow (\text{education}(x) \in \text{Higher education}) (GR=0.67)$

$V(x) = \text{Anies}(x)[3] - \text{ahok}(x)[2] \rightarrow (\text{education}(x) \in \text{Higher education}) (GR=1.5)$

Data mining if it is implemented then will be powerful for prediction purposes, where politician will be helpful to predict their future politic career based on previous data and finding data mining knowledge patten/rule such as characteristic, discriminant, classification, association and clustering.

For next research, it will be interested when we can find out the preference of voters to their candidate couples' profile and programs. What size of the preferences of voters in choosing their candidate couples' profile and programs, how many voters interested with their chosen candidate couple' profile and how many voters interested with their chosen candidate couple' programs. Meanwhile, we have never had the data about how many voters who voted based on profile or programs preference, since the laws protect the secrets of their ballots. However, we can have the voters' characteristic data.

REFERENCES

- [1] Mikut, R. and Reischl, M. 2011. Data mining tools. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(5), 431-445.
- [2] Piatetsky-shapiro, G. 2011. Algorithms for data analysis/data mining in 2011, retrieved March 6, 2012, from <http://www.kdnuggets.com/polls/2011/algorithms-analytics-data-mining.html>
- [3] Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J. and Steinberg, D. 2007. Top 10 algorithms in data mining. Knowl. Inf. Syst., 14(1), 1-37.
- [4] General Election Commissions Province of DKI Jakarta, Profile and Vision-Mission of Governor Candidate and Vice Governor of DKI Jakarta Year 2017, Access on 25 April 2017 from <https://kpujakarta.go.id/download/>
- [5] General Election Commissions, Hasil hitung TPS (Form C1) Provinsi DKI Jakarta Putaran ke 1, Access on 25 April 2017 from: https://pilkada2017.kpu.go.id/hasil/t1/dki_jakarta
- [6] General Election Commissions, Hasil hitung TPS (Form C1) Provinsi DKI Jakarta Putaran ke 2, Access on 25 April 2017 from: https://pilkada2017.kpu.go.id/hasil/2/t1/dki_jakarta
- [7] Utami, Y.T., Warnars, H.L.H.S. 2016. Penerapan Supervised Emerging Patterns Untuk Multi Atribut Pada Data Online Izin Usaha Pertambangan di Indonesia (Studi Kasus: EITI Indonesia). Jurnal Sistem Komputer, 6(2), 70-76.
- [8] Dong, G. Li, J. 1999. Efficient mining of emerging patterns: discovering trends and differences. In Proceedings of the Fifth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (San Diego, California, United States, August 15 - 18, 1999). KDD '99. ACM, New York, NY, 43-52.
- [9] Warnars, H.L.H.S. 2010. Attribute Oriented Induction with simple select SQL Statement. The 1st Int. Conf. on Computation for Science and Technology (ICCST-I), Chiang Mai, Thailand, 4-6 August 2010.
- [10] Warnars, H.L.H.S. 2010. Measuring Interesting rules in characteristic rule. The 2nd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIT), pp. 152-156, Bali, Indonesia, 1-2 July 2010.
- [11] Warnars, H.L.H.S. 2010. Classification rule with simple select SQL statement. National seminar Budi Luhur University 2010, Budi Luhur University, Jakarta, 5 August 2010.
- [12] Warnars, H.L.H.S. 2010. Attribute oriented induction with star schema. Int. Journal of Database Management system (IJDBMS), 2(2), 20-42.
- [13] Warnars, H.L.H.S. 2010. Star Schema Design for Concept Hierarchy in Attribute Oriented Induction. Internetworking Indonesia Journal, 2(2), 33-39.
- [14] Warnars, H.L.H.S. 2015. Mining Patterns with Attribute Oriented Induction. The Int. Conf. on Database, Data Warehouse, Data Mining and Big Data (DDMBD2015), Tangerang, Indonesia, pp. 11-21.
- [15] Warnars, H.L.H.S., Wijaya, M.I., Tjung, H.B., Xaverius, D.F., Hauten, D.V., Sasmoko. 2016. Easy understanding of Attribute Oriented Induction (AOI) characteristic rule algorithm. International journal of Applied Engineering Research (IAER), 11(8), 5369-5375.
- [16] Wibowo, A., Warnars, H.L.H.S. 2016. Pengembangan learning characteristic rule pada algoritma data mining attribute oriented induction. Jurnal Sistem Komputer, 6(1), 17-29.
- [17] Warnars, H.L.H.S. 2012. Attribute Oriented Induction of High-level Emerging Patterns. International Symposium on Foundations and Frontiers of Data Mining in conjunction with IEEE Int. Conf. on Granular Computing (IEEE GrC2012), Hangzhou, China, 11-13 August 2012.
- [18] Warnars, H.L.H.S. 2014. Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) future research. The 8th Int. Conf. on Information & Communication Technology and Systems (ICTS), Surabaya, Indonesia, pp. 13-18, 24-25 September 2014.
- [19] Warnars, H.L.H.S. 2014. Mining Frequent Pattern with Attribute Oriented Induction High level Emerging Pattern (AOI-HEP). The 2nd Int. Conf. on Information and Communication Technology (IEEE ICICT 2014), Bandung, Indonesia, pp. 144-149, 28-30 May 2014.
- [20] Warnars, H.L.H.S. 2014. Mining Frequent and Similar Patterns with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP) Data Mining Technique. Int. Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 3(11), 266-276.
- [21] Warnars, H.L.H.S. 2016. Using Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) to mine frequent patterns. Int. Journal of Electrical and Computer Engineering (IJECE), 6(6), 3037-3046.