Strengthening campus finance by analyzing attribute attributes for student registration classifications

M. Adib Al Karomi 1, Much. Rifqi Maulana 2, Slamet Joko Prasetiyono 3, Ivandari 4, Arochman 5

1-4 Computer Science, STMIK Widya Pratama Pekalongan, Indonesia

5 Computerized accounting, STMIK Widya Pratama Pekalongan, Indonesia

Abstract— Students are the most valuable assets in a private college. Assets like this that really need to be maintained and maintained, because most of the income from the private campus is derived from the tuition fees of students. The large number of students who resigned and did not conduct registration would have an impact on the financial institutions. STMIK Widya Pratama is the only computer science campus in Pekalongan City. Data from the last 5 years obtained from the new student admissions committee at STMIK Widya Pratama Pekalongan shows that out of 2670 prospective students who enroll, there are at least 514 prospective students who do not register. This means that around 20% of students do not register. Several analyzes related to the classification for student registration were conducted. In this case the best method that can be used is C45. In the process of calculating the C45 algorithm, information gain method is used to determine the importance of data attributes. The calculation results show that the attribute with the highest level of importance is the city\_district attribute from the prospective student's residence, followed by the attributes of education, parental education, and tuition. These results can later be used and developed to create a system to support campus policy.

Index Terms— student data attribute, information gain.

1. Introduction

College of Informatics and Computer Management (STMIK) Widya Pratama is the only computer college in Pekalongan City. STMIK Widya Pratama was established in 2002 based on Minister of National Education Decree Number: 149 / D / O / 2002. In 2004 an AMIK and STMIK merger was based on the Minister of National Education Decree Number: 75 / D / O / 2004. As a private tertiary institution, STMIK requires a lot of operational funds, one of the biggest supporters of operational funds is from students. The number of students is one of the benchmarks of private universities in general.

Every year STMIK Widya Pratama opens new student registration. Publications and promotions have been carried out by visiting high school / vocational school or equivalent to provide exposure related to facilities and general description of lectures at STMIK. In fact, every year there is always a significant difference in the number of registrants with the number of students who have registered. Table 1 is data obtained from the new student admission committee STMIK Widya Pratama Pekalongan.

The decline in the number of registered students can affect the main income of institutions, especially private universities. Several actions were taken to strive to increase the number of new students. one of them is by providing payment relief or providing scholarships to high-achieving students. The significant difference between the number of registrants and the number of students registering is one problem that must be solved immediately. The main institution of the new student admission team must try more to increase the number of students or at least maintain the existing number. If the possibility of prospective students who do not conduct heregistration can be known earlier, action can be taken to defend the prospective students [1].

Classification is one of the main functions of data mining [2]. The classification process can be done in various ways [3]. In the classification process various types of classification algorithms can be used [4]. In a previous study a comparison was made between several algorithms for STMIK Widya Pratama student registration classification. The results of the study prove that the C45 algorithm is the best for the prospective student classification. In the C45 algorithm stage weighting calculations are used for all dataset attributes. This study will perform calculations to determine the importance of all data attributes in the STMIK Widya Pratama student registration data. The results of this study can later be used as an analysis of the acceptance of new students at STMIK. The expectation of this analysis can be used as a strategy to increase the number of new students at STMIK Widya Pratama Pekalongan.

Table 1. List of student registrants and registrations

(source: PMB STMIK Widya Pratama Pekalongan 31-August-2018)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Year** | **All Registrants** | **Registration** | **Not Registration** | **Percentage** | |
| **Registration** | **Not Registration** |
| **2013** | 705 | 513 | 192 | 73% | 27% |
| **2014** | 608 | 474 | 134 | 78% | 22% |
| **2015** | 667 | 500 | 167 | 75% | 25% |
| **2016** | 654 | 427 | 227 | 77% | 23% |
| **2017** | 506 | 428 | 78 | 84,6% | 15,4% |
| **2018** | 401 | 270 | 131 | 67,3% | 32,7% |
|  |  |  |  |  |  |

2. Literature Review

2.1. Related Research

The research conducted by M. Adib Alkaromi [5] conducted a classification of prospective students by using a dataset of 2389 records. The dataset used is PMB data from 2011 to 2013. Attributes in the dataset are 44 with one label attribute and one id attribute. All attributes of feature selection using information gain are then classified using the K-NN algorithm. In the study the K-NN algorithm obtained an accuracy rate of 78.15% using all of its attributes. Whereas after feature selection using information gain accuracy of the K-NN algorithm rose to reach 83.93%. This study uses a configuration matrix for evaluation and 10olds cross validation for the validation process.

2.2. Data Mining

Data Mining or often also called Knowledge Discovery in Databases (KDD) is a field of science that mostly discusses the pattern of a data. A series of processes to obtain knowledge or patterns from a data set is called data mining [2]. Data mining is an extraction process to obtain information that was previously unknown from a data [2]. Data mining can analyze old cases to find patterns from data using pattern recognition techniques such as statistics and mathematics [3]. Large data collection can be meaningless if the value of information in it is lacking or can not be taken a knowledge .. Data mining answers this problem by analyzing large data then making a certain rule, pattern, or model to recognize new data that is not in row of data stored [6].

Data mining has several functions. The main functions of data mining are: Estimation, Prediction, Clustering, Classification, and Associations. Based on the learning method, the data mining function is divided into 2 [7], namely Supervised Learning, Unsupervised Learning. Supervised learning must have sample data or often referred to as training data. Whereas in unsupervised learning does not require training data. Classification is a data mining function that is classified as supervised learning.

2.2. Information gain

Information gain is a feature selection method that is widely used by researchers [8]. Using information gain limits from the importance of an attribute can be determined [9]. The information gain value is the entropy value before the separation is reduced by the entropy value after separation. This value measurement is used for weighting attributes which will later be used in the classification stage or not. Attributes that meet certain weighting criteria will be used in the process of classifying an algorithm.

In general, information gain stages are carried out in the following ways:

1. Calculates the information gain value for all attributes in the original dataset.
2. Determine the desired threshold. The existence of this limitation allows attributes that are weighted equal to the limit or greater than the boundary to be maintained and used in the classification stage. Then attributes with a weight below the limit will be discarded and not used in the classification stage.
3. The dataset is updated using only the selected attribute.

3. Research Methods

The stages in this study can be seen as shown in figure 1 below

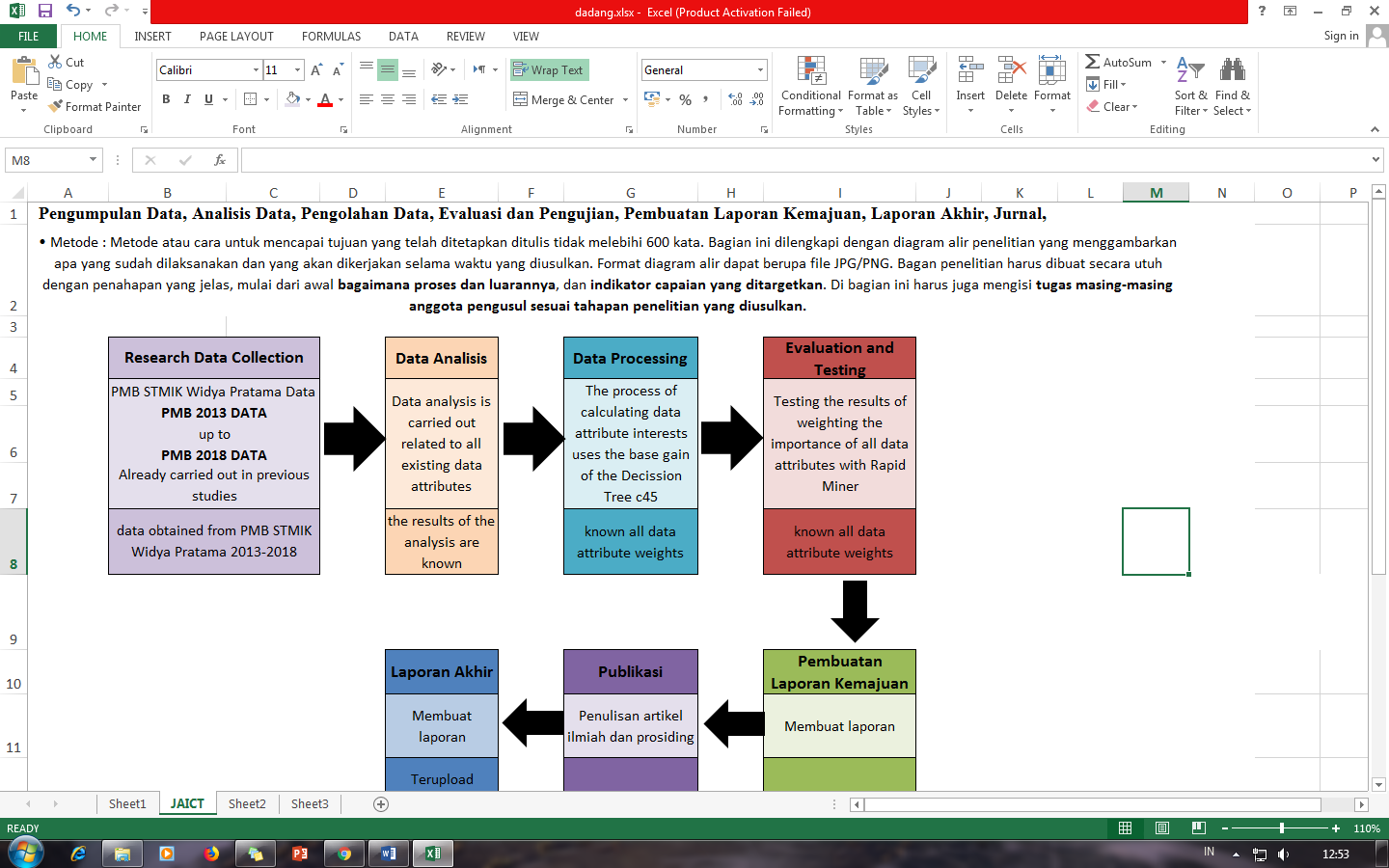


Fig.1 Stages of research

In more detail the stages of research carried out are as follows:

3.1. Data Collection Method

Data collection is done by taking data from the new student admission committee STMIK Widya Pratama Pekalongan. PMB data from 2013 to 2017 has been obtained and used as prior research material. The PMB data for 2018 is still developing because on August 30 this new student admission process is still ongoing and will only be closed on September 21. From the data obtained, there are 43 data attributes with one objective attribute, registration. The data obtained has been in the form of a file with the xlsx extension which is the result of downloading from the PMB STMIK Widya Pratama Pekalongan application.

3. 2. Data Analysis

Stages of data analysis are done to narrow down the existing data attributes. Of the 43 data attributes, there are attributes with very many variants such as KTP numbers. Each record has a different KTP number. In addition there are address attributes and names and names of parents with conditions that are almost the same. Some of the attributes as mentioned above will not be used in the next process. In the process of data analysis, there will also be repairs to some irrelevant data entries or errors filled in by the committee when inputting data.

3. 3. Data Processing

The processing stage will be carried out using all existing data, namely PMB data from 2013 to 2018. The processing process is the core process of this research. This process uses the help of a rapid miner application for data processing. In addition, the information gain algorithm is used to determine the importance of all existing attributes. The final result of this stage is to know the importance of all existing attributes.

4. Results and Discussion

3. 3. Data obtained

As explained earlier the data used is student registration data at STMIK Widya Pratama Pekalongan in the last 5 years. Table 2 is the metadata of the dataset that has been obtained:

Table 2. Student Heregistration Data Metadata

(source: PMB STMIK Widya Pratama Pekalongan 31-August-2018)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ROLE** | **NAME** | **TYPE** | **STATISTICS** | **RANGE** | **MISSINGS** |
| label | registrasi | binominal | mode = Ya (2152), least = Tidak (514) | Tidak (514), Ya (2152) | 0 |
| regular | no\_daftar | polynominal | mode = ~ | ~ | 0 |
| regular | nama | polynominal | mode = ~ | ~ | 0 |
| regular | alamat | polynominal | mode = ~ | ~ | 0 |
| regular | kota\_kec | polynominal | mode = PEKALONGAN (1342), least = BANYUMAS (1) | PEKALONGAN (1342), BATANG (578), PEMALANG (294), KAB PEKALONGAN (221), PEKALONGAN KABUPATEN (64), KAB BATANG (35), KAB PEMALANG (33), TEGAL (15), BATANG KABUPATEN (11), KEDUNGWUNI (7), WIRADESA (7), ACEH (5), BREBES (4), AMBON (3), BANJARNEGARA (3), KARANGANYAR (3), BOYOLALI (2), COMAL (2), DEMAK (2), JAKARTA (2), KENDAL (2), SEMARANG (2), YOGYAKARTA (2), BANDAR (1), BANDUNG (1), BANYUMAS (1), ... and 4 more ... , BOGOR (1), CURUP (1), DEPOK (1), DORO (1), JEPARA (1), KALIMANTAN TIMUR (1), KLATEN (1), KUDUS (1), LAMPUNG (1), MAGELANG (1), MANADO (1), PADANG PANJANG (1), PALEMBANG (1), PATI (1), PEMALANG KABUPATEN (1), PONTIANAK (1), PURBALINGGA (1), REMBANG (1), SEKADAU (1), SOLO (1), SUMATRA UTARA (1), SURABAYA (1), TEGAL KABUPATEN (1), WONOGIRI (1), WONOKERTO (1), WONOSOBO (1) | 0 |
| regular | Kode\_pos | polynominal | mode = ~ | ~ | 0 |
| regular | Telp | polynominal | mode = ~ | ~ | 0 |
| regular | tmplhr | polynominal | mode = ~ | ~ | 0 |
| regular | Tglhr | polynominal | mode = ~ | ~ | 0 |
| regular | Jenjang | binominal | mode = S1 (2176), least = D3 (494) | D3 (494), S1 (2176) | 0 |
| regular | Jenjang2 | integer | mode = S1 (814), least = D3 (523) | D3 (523), S1 (814) | 0 |
| regular | Kelas | polynominal | mode = REGULER (2537), least = TRANSFER (30) | REGULER (2537), ALIH JENJANG (103), TRANSFER (30) | 0 |
| regular | prodi | polynominal | avg = 215.955 +/- 47.255 | [110.000 ; 240.000] | 0 |
| regular | prodi2 | polynominal | avg = 92.404 +/- 101.084 | [0.000 ; 240.000] | 0 |
| regular | agama | polynominal | mode = ISLAM (2582), least = KONGHUCU (1) | ISLAM (2582), KATHOLIK (43), KRISTEN (42), KONGHUCU (1), HINDU (2) | 0 |
| regular | st\_kerja | polynominal | mode = BELUM BEKERJA (2059), least = BEKERJA (545) | BELUM BEKERJA (2059), BEKERJA (545) | 0 |
| regular | jnkel | binominal | mode = LAKI-LAKI (1737), least = PEREMPUAN (918) | LAKI-LAKI (1737), PEREMPUAN (918) | 14 |
| regular | warganeg | binominal | mode = WNI (2664), least = null (4) | WNI (2664), null (4) | 2 |
| regular | stssipil | binominal | mode = TIDAK MENIKAH (2589), least = MENIKAH (53) | TIDAK MENIKAH (2589), MENIKAH (53) | 20 |
| regular | aslskl | polynominal | mode = ~ | ~ | 0 |
| regular | KOTAASLSKL | polynominal | mode = ~ | ~ | 0 |
| regular | th\_lulus | integer | avg = 2007.071 +/- 116.772 | [0.000 ; 2017.000] | 0 |
| regular | no\_ijasah | polynominal | mode = ~ | ~ | 1 |
| regular | nmortu | polynominal | mode = ~ | ~ | 0 |
| regular | almortu | polynominal | mode = ~ | ~ | 0 |
| regular | telportu | integer | mode = ~ | ~ | 0 |
| regular | kota\_kec\_ortu | polynominal | mode = ~ | ~ | 0 |
| regular | kodepos\_ortu | polynominal | mode = ~ | ~ | 0 |
| regular | krjortu | polynominal | mode = LAIN-LAIN (174), least = PETANI (1) | PNS (86), SWASTA (66), WIRAUSAHA (164), LAIN-LAIN (174), POLRI/TNI (13), BURUH (2), PETANI (1) | 0 |
| regular | INSTANSI\_ORTU | binominal | mode = ~ | ~ | 141 |
| regular | PANGKAT\_ORTU | binominal | mode = ~ | ~ | 140 |
| regular | NIP\_ORTU | polynominal | mode = ~ | ~ | 506 |
| regular | pend\_ortu | polynominal | mode = SMA (942), least = D2 (1) | SD (774), SMA (942), SMP (475), TIDAK SEKOLAH (100), S1 (304), D3 (44), S2 (26), D2 (1), D1 (3) | 0 |
| regular | info | polynominal | mode = BROSUR (2429), least = SEKOLAH (11) | TEMAN (121), BROSUR (2429), TIM MARKETING (63), SEKOLAH (11), INTERNET (45) | 0 |
| regular | kd\_daftar | binominal | mode = SENDIRI (2601), least = 0 (1) | SENDIRI (2601), MARKETING MAHASISWA STMIK (41), MARKETING SEKOLAH (26), 0 (1) | 1 |
| regular | ST\_TEST | binominal | mode = TEST (1884), least = BEBAS TEST (784) | TEST (1884), BEBAS TEST (784) | 1 |
| regular | jam | polynominal | mode = ~ | ~ | 0 |
| regular | tgl\_daftar | polynominal | mode = ~ | ~ | 0 |
| regular | kd\_user | polynominal | mode = ~ | ~ | 0 |
| regular | marketing | polynominal | mode = ~ | ~ | 505 |
| regular | gelombang | polynominal | mode = 2 (1226), least = 0 (6) | 1 (873), YYS-01 (120), 2 (1226), 3 (434), 0 (6), YYS-99 (10) | 0 |
| regular | shifkelas | binominal | mode = PAGI (1906), least = MALAM (762) | PAGI (1906), MALAM (762) | 0 |
| regular | biaya | integer | avg = 150000 +/- 0 | [150000.000 ; 150000.000] | 0 |
| regular | gelgra | polynominal | mode = ~ | ~ | 0 |
| regular | PEKERJAAN | polynominal | mode = BELUM BEKERJA (1656), least = POLRI/TNI (1) | BELUM BEKERJA (1656), SWASTA (148), PNS (641), WIRAUSAHA (50), LAIN-LAIN (173), POLRI/TNI (1) | 0 |
| regular | ALMTKERJA | polynominal | mode = ~ | ~ | 407 |
| regular | JBTNKERJA | binominal | mode = ~ | ~ | 419 |
| regular | TELPKERJA | polynominal | mode = ~ | ~ | 410 |
| regular | biayakuliah | polynominal | mode = ORANG TUA / WALI (1569), least = BEASISWA RAPOR >=8,5 (107) | SENDIRI (363), ORANG TUA / WALI (1569), BEASISWA RANGKING 10 BESAR (357), BEASISWA YAYASAN (129), BEASISWA PRESTASI (144), BEASISWA RAPOR >=8,5 (107) | 0 |
| regular | KD\_KONSENTRASI | integer | avg = 21595.988 +/- 4726.259 | [11001.000 ; 24002.000] | 0 |
| regular | KD\_KONSENTRASI2 | integer | avg = 9244.549 +/- 10109.355 | [0.000 ; 24002.000] | 0 |
| regular | ukuranalmamater | polynominal | mode = L (234), least = LLL (1) | M (153), XL (98), L (234), 3L (15), LLL (1), S (3), - (2) | 0 |
| regular | REGISTER | integer | avg = 0.846 +/- 0.361 | [0.000 ; 1.000] | 0 |
| regular | DIPINDAHKAN | binominal | mode = no (506), least = no (506) | no (2668) | 0 |
| regular | JENJANGDIPINDAHKAN | binominal | mode = no (506), least = no (506) | no (2668) | 0 |
| regular | KETLULUS | binominal | mode = lls (2668), least = tls (1) | tls (1), lls (2668) | 0 |
| regular | FB | polynominal | mode = ~ | ~ | 105 |
| regular | BB | polynominal | mode = ~ | ~ | 107 |
| regular | no\_online | integer | mode = ~ | ~ | 2 |
| regular | NIK | real | mode = ~ | ~ | 0 |

4.2. Data analysis

From the data obtained there are several missing values and attributes with unlimited variants. This attribute obviously will not be used in the calculation process because there are many types equal to the number of records that are available. This study only uses attributes that are considered relevant. Manually selecting attributes is done by looking at the type of attribute and the number of variants of the attribute. Table 3 shows the attributes selected and used in the next process.

Table 3 Student Heregistration Data Metadata

(source: PMB STMIK Widya Pratama Pekalongan 31-August-2018)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Role | Name | Type | Statistics | Range | Missings |
| id | no\_daftar | polynominal | mode = 13-010-0001 (1), least = 13-010-0001 (1) | 13-010-0001 (1), 13-010-0005 (1), 13-010-0006 (1), 13-010-0027 (1), 13-010-0035 (1), 13-010-0046 (1), 13-010-0065 (1), 13-010-0066 (1), 13-010-0067 (1), 13-010-0083 (1), 13-010-0084 (1), 13-010-0085 (1), 13-010-0092 (1), 13-010-0093 (1), 13-010-0094 (1), 13-010-0095 (1), 13-010-0118 (1), 13-010-0122 (1), 13-010-0126 (1), 13-010-0181 (1), 13-010-0190 (1), 13-010-0191 (1), 13-010-0194 (1), 13-010-0195 (1), 13-010-0197 (1), 13-010-0206 (1), ... and 2620 more ... , 17-040-0430 (1), 17-040-0431 (1), 17-040-0434 (1), 17-040-0435 (1), 17-040-0436 (1), 17-040-0438 (1), 17-040-0439 (1), 17-040-0443 (1), 17-040-0444 (1), 17-040-0448 (1), 17-040-0452 (1), 17-040-0454 (1), 17-040-0457 (1), 17-040-0467 (1), 17-040-0470 (1), 17-040-0471 (1), 17-040-0473 (1), 17-040-0474 (1), 17-040-0475 (1), 17-040-0476 (1), 17-040-0494 (1), 17-040-0496 (1), 17-040-0498 (1), 17-040-0500 (1), 17-040-0502 (1), 17-040-0504 (1) | 0 |
| label | registrasi | binominal | mode = Ya (2152), least = Tidak (514) | Tidak (514), Ya (2152) | 4 |
| regular | kota\_kec | polynominal | mode = PEKALONGAN (1342), least = BANYUMAS (1) | PEKALONGAN (1342), BATANG (578), PEMALANG (294), KAB PEKALONGAN (221), PEKALONGAN KABUPATEN (64), KAB BATANG (35), KAB PEMALANG (33), TEGAL (15), BATANG KABUPATEN (11), KEDUNGWUNI (7), WIRADESA (7), ACEH (5), BREBES (4), AMBON (3), BANJARNEGARA (3), KARANGANYAR (3), BOYOLALI (2), COMAL (2), DEMAK (2), JAKARTA (2), KENDAL (2), SEMARANG (2), YOGYAKARTA (2), BANDAR (1), BANDUNG (1), BANYUMAS (1), ... and 4 more ... , BOGOR (1), CURUP (1), DEPOK (1), DORO (1), JEPARA (1), KALIMANTAN TIMUR (1), KLATEN (1), KUDUS (1), LAMPUNG (1), MAGELANG (1), MANADO (1), PADANG PANJANG (1), PALEMBANG (1), PATI (1), PEMALANG KABUPATEN (1), PONTIANAK (1), PURBALINGGA (1), REMBANG (1), SEKADAU (1), SOLO (1), SUMATRA UTARA (1), SURABAYA (1), TEGAL KABUPATEN (1), WONOGIRI (1), WONOKERTO (1), WONOSOBO (1) | 0 |
| regular | Jenjang | binominal | mode = S1 (2176), least = D3 (494) | D3 (494), S1 (2176) | 0 |
| regular | Jenjang2 | binominal | mode = S1 (814), least = D3 (523) | D3 (523), S1 (814) | 1333 |
| regular | Kelas | polynominal | mode = REGULER (2537), least = TRANSFER (30) | REGULER (2537), ALIH JENJANG (103), TRANSFER (30) | 0 |
| regular | prodi | integer | avg = 215.955 +/- 47.255 | [110.000 ; 240.000] | 0 |
| regular | prodi2 | integer | avg = 92.404 +/- 101.084 | [0.000 ; 240.000] | 0 |
| regular | agama | polynominal | mode = ISLAM (2582), least = KONGHUCU (1) | ISLAM (2582), KATHOLIK (43), KRISTEN (42), KONGHUCU (1), HINDU (2) | 0 |
| regular | st\_kerja | binominal | mode = BELUM BEKERJA (2059), least = BEKERJA (545) | BELUM BEKERJA (2059), BEKERJA (545) | 66 |
| regular | jnkel | binominal | mode = LAKI-LAKI (1737), least = PEREMPUAN (918) | LAKI-LAKI (1737), PEREMPUAN (918) | 15 |
| regular | warganeg | binominal | mode = WNI (2664), least = null (4) | WNI (2664), null (4) | 2 |
| regular | stssipil | binominal | mode = TIDAK MENIKAH (2589), least = MENIKAH (53) | TIDAK MENIKAH (2589), MENIKAH (53) | 28 |
| regular | th\_lulus | integer | avg = 2007.071 +/- 116.772 | [0.000 ; 2017.000] | 0 |
| regular | Fresh Graduate | integer | avg = 0.879 +/- 2.037 | [0.000 ; 24.000] | 9 |
| regular | pend\_ortu | polynominal | mode = SMA (942), least = D2 (1) | SD (774), SMA (942), SMP (475), TIDAK SEKOLAH (100), S1 (304), D3 (44), S2 (26), D2 (1), D1 (3) | 1 |
| regular | info | polynominal | mode = BROSUR (2429), least = SEKOLAH (11) | TEMAN (121), BROSUR (2429), TIM MARKETING (63), SEKOLAH (11), INTERNET (45) | 1 |
| regular | kd\_daftar | polynominal | mode = SENDIRI (2601), least = 0 (1) | SENDIRI (2601), MARKETING MAHASISWA STMIK (41), MARKETING SEKOLAH (26), 0 (1) | 1 |
| regular | ST\_TEST | binominal | mode = TEST (1884), least = BEBAS TEST (784) | TEST (1884), BEBAS TEST (784) | 2 |
| regular | gelombang | polynominal | mode = 2 (1226), least = 0 (6) | 1 (873), YYS-01 (120), 2 (1226), 3 (434), 0 (6), YYS-99 (10) | 1 |
| regular | shifkelas | binominal | mode = PAGI (1906), least = MALAM (762) | PAGI (1906), MALAM (762) | 2 |
| regular | PEKERJAAN | polynominal | mode = BELUM BEKERJA (1656), least = POLRI/TNI (1) | BELUM BEKERJA (1656), SWASTA (148), PNS (641), WIRAUSAHA (50), LAIN-LAIN (173), POLRI/TNI (1) | 1 |
| regular | biayakuliah | polynominal | mode = ORANG TUA / WALI (1569), least = BEASISWA RAPOR >=8,5 (107) | SENDIRI (363), ORANG TUA / WALI (1569), BEASISWA RANGKING 10 BESAR (357), BEASISWA YAYASAN (129), BEASISWA PRESTASI (144), BEASISWA RAPOR >=8,5 (107) | 1 |
| regular | KD\_KONSENTRASI | integer | avg = 21595.988 +/- 4726.259 | [11001.000 ; 24002.000] | 1 |
| regular | KD\_KONSENTRASI2 | integer | avg = 9244.549 +/- 10109.355 | [0.000 ; 24002.000] | 1 |
| regular | KETLULUS | binominal | mode = lls (2668), least = tls (1) | tls (1), lls (2668) | 1 |

4.3. The results of weighting the importance of data attributes

Calculations using information gain are done with the help of the rapid miner application. The process of calculation using the rapid miner produces weights for each different attribute. Table 4 is the result of calculating information gain.

Table 4 Results of weighting for each attribute

(Processed with rapid miner)

|  |  |
| --- | --- |
| **Atribute** | **Weight** |
| warganeg | 0.0 |
| Kelas | 0.0025305405258738184 |
| st\_kerja | 0.009296169826497223 |
| stssipil | 0.030387921658118308 |
| Fresh Graduate | 0.04655467177039355 |
| KETLULUS | 0.059488060174902435 |
| PEKERJAAN | 0.06056195746833359 |
| shifkelas | 0.07868769232845471 |
| info | 0.07880154664430117 |
| Jenjang | 0.0825337195905499 |
| prodi | 0.12088559759218581 |
| KD\_KONSENTRASI | 0.12088559759218581 |
| ST\_TEST | 0.18377378063464517 |
| th\_lulus | 0.24897354370989158 |
| agama | 0.2726527521100228 |
| kd\_daftar | 0.30439808333395335 |
| prodi2 | 0.33769481477402047 |
| KD\_KONSENTRASI2 | 0.35042128421087254 |
| gelombang | 0.447795250741339 |
| jnkel | 0.46871080716174796 |
| biayakuliah | 0.541072325203167 |
| pend\_ortu | 0.5521100988478882 |
| Jenjang2 | 0.7295855407801191 |
| kota\_kec | 1.0 |

4.4. Discussion

Table 4 is the result of research that can be interpreted: (1). The city attribute of the sub-district is the attribute with the highest information gain value, namely 1. It means that the city attribute of the sub-district is the attribute that most influences student registration. (2). The citizens attribute has an information gain of 0. It means that the attributes of citizens do not affect the classification of student registration at all. This happens because all prospective students have the same citizenship, namely Indonesian citizens. (3). The interests of other attributes can be seen as table 4. The top order is an attribute that is less influential in accordance with its value. While the bottom is the most influential attribute with 1.

5. Conclusion

From the results of the study it can be concluded that the most dominant attribute that influences student registration is the city / sub-district attribute, level, parental education, tuition fees and so on. This knowledge can be used as a reference for the campus management to increase the number of students who register by applying certain policies. The existence of this knowledge if supported by the right policy will reduce the process of students who are resigning or not registering.

6. Acknowledgment

This research is part of a research funded by DIKTI in a Beginners Lecturer Research Scheme for fiscal year 2019.

References

[1] Kusrini, & Taufiq, L. E. (2009). Algoritma Data Mining. Yogyakarta: Andi Offset.

[2] Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition. Elsevier.

[3] Larose, D. T. (2005). Discovering Knowledge in Data: an Introduction to Data Mining. John Wiley & Sons.

[4] Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques Second Edition. Elsevier. Elsevier.

[5] Alkaromi, M. A. (2014). Information Gain untuk Pemilihan Fitur pada Klasifikasi Heregistrasi Calon Mahasiswa dengan Menggunakan K-NN.

[6] Prasetyo, E. (2012) Data Mining Konsep dan Aplikasi menggunakan Matlab. Yogyakarta: Andi Offset.

[7] Santosa, B. (2007) Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Edisi Pert. Yogyakarta: Graha Ilmu.

[8] Deng, H. and Runger, G. (2012) ‘Feature Selection via Regularized Trees’. Available at: http://arxiv.org/abs/1201.1587v3 (Accessed: 16 October 2014).

[9] Novakovic, J. (2010) ‘The Impact of Feature Selection on the Accuracy of 1DwYH Bayes Classifier’, 2, pp. 1113–1116.