

Classification Of Category Selection Title Undergraduate Thesis Using K-Nearest Neighbor Method

Ratih Kumalasari Niswatin^{*}, Ardi Sanjaya

Departement of Informatics Engineering/University of Nusantara PGRI Kediri, Campus II, Mojoroto
Gang I No. 6, Mojoroto, Kediri, Jawa Timur, Indonesia

^{*}Corresponding author, e-mail: ratih.workmail@gmail.com

Abstract

This research makes the classification system of category selection title undergraduate thesis title use k-nearest neighbor method. This research will be conducted on the students of Informatics Engineering Department Faculty of Engineering, Universitas Nusantara PGRI Kediri. The purpose of making this system is to employee department and students to more easily make a classification of category selection undergraduate thesis title based on the field of interest and field of expertise of each student. The method used to classify the selection of undergraduate thesis title categories is k-nearest neighbor method using several criteria based on students' interests and expertise in a particular field or course. The result of this sitem is an information category of undergraduate thesis title of students who have been processed based on the field of interest and field of expertise of each student.

Keywords: classification, undergraduate thesis, k-nearest neighbor

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

The university is the organizer of academic education for students [1]. Students as college products can be used as a reference to show the success of education. The success of student achievement can be seen from several factors such as cumulative grade point value (GPA) obtained by students, duration of student study, and student thesis score.

In this research will be discussed one of the above factors is the success of student undergraduate thesis score. Understanding undergraduate thesis according to a large dictionary of the Indonesian language is a scientific essay that must be written by students as part of the final requirements of academic education [2]. Based on this, every final year student must take and complete the undergraduate thesis as the graduation requirement from higher education. But in reality many students have difficulty in choosing the category and title of the graduate thesis according to the field of interest and the field of expertise, therefore in this research will be made a classification system of category selection of student undergraduate thesis title using k-nearest neighbor (k-nn) method.

Nearest neighbor is an approach to finding a case by calculating the proximity between a new case and an old case, based on a weighted fit of a number of features [3]. Nusantara University PGRI Kediri is a private university in Kediri City which has five Faculties, one of which is the Faculty of Engineering which has five study programs namely: Informatics Engineering, Mechanical Engineering, Electrical Engineering, Industrial Engineering and Information Systems. This research will be conducted in Informatics Engineering Study Program to assist the department in classifying the category of student's undergraduate thesis title and assisting the final grade students to choose the title category of undergraduate thesis based on the field of interest and the area of expertise.

The problems can be formulated in this research are how to apply k-nearest neighbor method to classify the selection of undergraduate thesis title categories of students based on their field of interest and field of expertise and how to apply k-nearest neighbor method for the proper and accurate classification of category selection of student undergraduate thesis title. The purpose of this research is to Apply k-nearest neighbor method for classification of undergraduate thesis title categories based on the field of interest and the field of expertise of

each student in Informatics Engineering Program Faculty of Engineering, Universitas Nusantara PGRI Kediri.

Benefits and contributions of this research are assisting the employee department of informatics to classify the category of student undergraduate thesis title in Informatics Engineering Program Faculty of Engineering Universitas Nusantara PGRI Kediri and assisting the final grade students to select undergraduate thesis title categories based on the areas of interest and the areas of expertise. Some of the previous studies that underlie this research are: Research conducted by Kiki Rizky Ananda S in 2014 with the title Decision Support System To Determine Title Thesis Department of Computer Information Techniques Using Simple Additive Weighthing Method (SAW). This research makes a decision support system to help the students to determine the desired thesis title by using simple additive weighing method [4].

Research conducted by Ratih Kumalasari Niswatin in 2015 under the title Decision Support System New Student Placement Using K-Nearest Neighbor Method. This study discusses the web-based decision support system to provide recommendation placement of new student majors are advised to enter the majors of informatics engineering or information systems using k-nearest neighbor method [5]. Research conducted by Khafiizh Hastuti in 2012 with the title of Comparative Analysis of Data Mining Classification Algorithm for Non-Active Student Prediction. This study comparable some data mining classification algorithms for predictions of non-active students, the goal is to determine the accuracy of each algorithm. The algorithm compiled is logistic regression, decision tree, naive bayes and neural network [6].

Definition of decision support system, according to Alter decision support system is an interactive information system that provides information, modeling, and manipulation of data used to assist decision making in semi-structured situations and unstructured situations where no one knows for sure how decisions should be made [7]. Definition of data mining, data mining is a term used to describe the discovery of knowledge in the database. Data mining and knowledge discovery in databases (KDD) are often used interchangeably to explain the process of extracting information in a large database. Data mining techniques are divided into several groups based on the tasks that can be done that is description, estimation, prediction, classification, clustering, association [3].

Classification is the process of finding a model or function that describes and distinguishes a data class or concept that aims to be used to predict the class of objects whose class labels are unknown. Data classification consists of two process steps, the first is the learning process (training phase) where the classification algorithm is made to analyze the training data and then represented in the form of classification rule, the second process is the classification where the test data is used to estimate the accuracy of the classification rule [8]. The classification process is based on four components, they are [9]: Class is a dependent variable in the form of categorical which represents the label contained in the object. Predictor is an independent variable that is represented by the characteristics of the data attribute. Training dataset is a set of data that contains the values of the two above components that are used to determine a suitable class based on predictors. Testing datasets are new data that will be classified by the model that has been created and the accuracy of the classification evaluated.

There are several algorithms that are often used in the classification technique that is k-nearest neighbor classification algorithm, decision tree, naive bayesian classification, and support vector machines. Nearest neighbor is an approach to finding a case by calculating the proximity between a new case and an old case, based on a weighted fit of a number of features [3]. The purpose of the nearest neighbor algorithm is to classify new objects based on attributes and training samples [10]. K-Nearest Neighbor is one of the methods used in classification.

The working principle of K-Nearest Neighbor is to find the closest distance between the data to be evaluated with the nearest neighbor in the training data. Here is a distance search formula using euclidian formula [11].

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Explanation :

x1 = data training

dist = distance

x2 = data testing
i = data variable

p = data dimension

2. Research Method

The method used for system development will be created using waterfall method concept. The following are the stages of the research method to be performed.

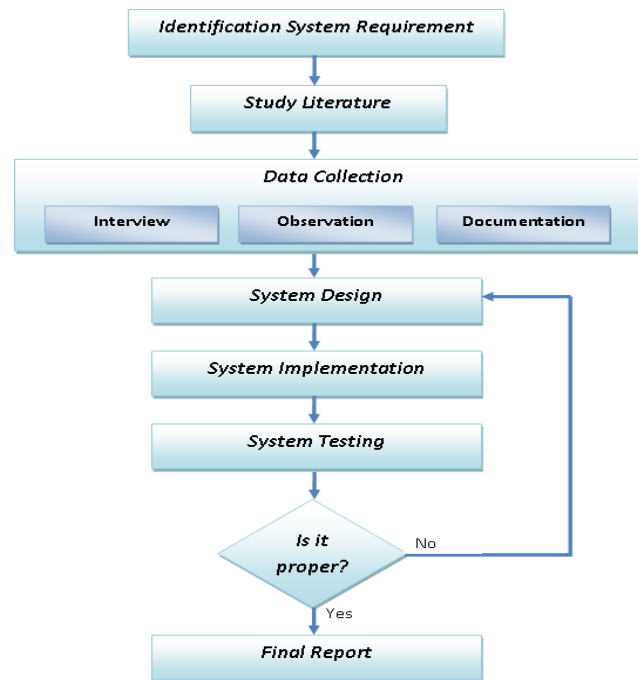


Figure 1. Stages Methods

Figure 1 the stages of research methods broadly described as follows:

- 1) Identification System Requirements
At this stage, the analysis of the needs of both software and hardware systems.
- 2) Study Literature
At this stage, the process of extracting information and study materials relating to research to be done, the materials studied are sourced from the relevant journals and books related to the research.
- 3) Data Collection
In this phase also carried out the data collection process by conducting interviews, observation and documentation of the data - the student data is needed. This stage will produce a document user requirements or data relating to the wishes of the user in the manufacturing system. This document will be the reference in the design stage of the system.
- 4) System Design
System design requirements will translate requirements into a software design before coding is made. In this process created the software architecture design, data structures, interface representations, and procedural algorithms.
- 5) Implementation
At this stage of the design process of translation into the language that can be recognized by the computer. In this process do the programming (coding) in accordance with the system. The programming language used in this system is PHP, the database used to store data is MySQL.

- 6) System Testing
In this stage testing of the program that was created by conducting tests on all functions and modules in the system.
- 7) Final Report
At this stage, the results of the report creation system manufacture.

3. Results and Analysis

In this section consists of two processes, they are explanation of k-nearest neighbor method and system design.

3.1. K-Nearest Neighbor Method

The classification system of category selection of student thesis title in this research is made using k-nearest neighbor method. The k-nearest neighbor method is chosen because it is the appropriate classification method used to classify the category of thesis titles based on the students' values on certain skill courses and based on the interest of the students in certain categories of fields. The criteria used in the classification of category selection of student undergraduate thesis title in this research is the value of information technology infrastructure planning course (PITI), the value of the software engineering course (RPL), the value of the network course, the value of the data mining course, the course processing grade Image, programming algorithm course values (ALPRO), database course values, and student interest in certain categories. Here are the steps - classification of category selection undergraduate thesis title students using k-nearest neighbor method.

1. Determining Data Training

Data training use student data force 2012 Informatics Engineering Program. Sample training data shown in table 1 below amounted to twenty student data. The attributes used are the value of information technology infrastructure planning course (PITI), the value of the software engineering course (RPL), the value of the network course, the data mining course grade, the value of the image processing course, the programming algorithm course (ALPRO) , The value of the database course, and student interest. The interest category attribute of the thesis title is converted into numbers with the conversion of number one for the interest of the information system category, the number two for the interest of the software engineering category, the number three for the interest of the network category, the number four for the interest of the data mining category, the number five for interest in the image category. In the training data is also displayed the results of the title category thesis students. Table 1 below is sample training data used.

Table 1. Data Training

NO	ID	NAMA	PITI	RPL	Network	Data Mining	Citra	Alpro	Databas e	Interest	category
1	12.1.03.02.0082	Adhen Bagus Putro U.	80	76	77	79	85	79	79	5	Citra
2	12.1.03.02.0203	Aditya Bayu Saputro	91	77	85	79	85	81	84	1	SI
3	12.1.03.02.0428	Ahmad Basyar Zakia	80	77	79	79	80	80	84	5	Citra
4	12.1.03.02.0208	Aldila Anggun Sasmita	70	80	79	79	80	81	84	3	Jaringan
5	12.1.03.02.0034	Andika Susanto	82	85	80	83	70	81	84	2	RPL
6	12.1.03.02.0035	Andri Eko Umarudin	62	80	77	85	60	80	81	4	Mining
7	12.1.03.02.0262	Angga Aditya Sardi	70	83	81	83	70	82	84	2	RPL
8	12.1.03.02.0062	Anggara Lexa P.	79	88	80	83	75	81	81	1	RPL
9	12.1.03.02.0227	Anggie Wibowo	69	80	77	80	70	80	80	2	RPL
10	12.1.03.02.0182	April Lia Tri Winarsih	91	78	80	81	85	81	86	1	SI
11	12.1.03.02.0429	Aprilina Pratiwi	88	82	79	75	85	80	78	1	SI
12	12.1.03.02.0367	Ari Nugroho	88	76	80	77	88	81	82	5	Citra
13	12.1.03.02.0073	Astriani Agus S.	91	83	80	83	85	81	81	1	SI

NO	ID	NAMA	PITI	RPL	Network	Data Mining	Citra	Alpro	Databas e	Interest	kategory
14	12.1.03.02.0395	Aviv Danar Putra P.	78	80	80	75	70	81	81	3	Jaringan
15	12.1.03.02.0136	Azis Zakaria	60	76	79	82	60	82	81	4	Mining
16	12.1.03.02.0321	Bayu Nur Rozikin	77	85	77	77	70	80	79	2	RPL
17	12.1.03.02.0210	Bayu Putra Prasetya	65	76	80	74	60	80	83	3	Jaringan
18	12.1.03.02.0122	Debit Bagus Dwicahya	68	83	80	73	70	80	82	2	RPL
19	12.1.03.02.0274	Delli Widianna Putri	74	81	79	73	80	82	81	5	Citra
20	12.1.03.02.0187	Deni Naviga A.	79	80	84	80	75	82	80	4	Mining

2. Determining Data Testing

Data testing is the data used as a system test, the classification of category selection of student's undergraduate thesis title is the data of the students of the 2013 Engineering Informatics Study Program.

Table 2. Data Testing

NO	ID	NAMA	PITI	RPL	Network	Data Mining	Citra	Alpro	Database	Interest
1	13.1.03.02.0347	Achmad Misbakhul H.	80	78	82	80	88	80	81	5
2	13.1.03.02.0026	Agung Pramuji	90	76	81	88	77	82	84	1
3	13.1.03.02.0222	Ahmad Eka Prasetya	71	76	80	80	76	81	81	3
4	13.1.03.02.0134	Ama Nilasari	90	80	77	80	78	80	80	1
5	13.1.03.02.0224	Arif Agus Triono	91	78	80	65	78	81	86	3
6	13.1.03.02.0220	Arto Dwiantoro	80	80	79	70	76	80	78	2
7	13.1.03.02.0180	Bagus Yayang F.	87	76	80	60	77	81	82	3
8	13.1.03.02.0042	Danang Permadi	92	76	77	75	77	80	79	1
9	13.1.03.02.0014	Defit Riza Nur Fadilah	90	76	80	75	81	80	83	5
10	13.1.03.02.0199	Destian Tri Handoko	77	76	80	90	83	80	82	4
11	13.1.03.02.0024	Fabbio Egi Iswahyudhi	88	75	79	85	76	80	78	4
12	13.1.03.02.0221	Fajar Mawardi	73	76	80	85	77	81	82	3
13	13.1.03.02.0111	Fuad Maftuhin	90	76	77	90	77	80	79	4
14	13.1.03.02.0230	Imam Nawawi	70	76	80	75	81	80	83	5
15	13.1.03.02.0065	M.Herma.Pradipta	70	76	80	60	77	81	82	3
16	13.1.03.02.0357	Miftahul Husna	93	76	80	70	77	81	82	1
17	13.1.03.02.0027	Mochamad Mahgribil H.	85	76	77	80	77	80	79	4
18	13.1.03.02.0028	Mohammad Fandi	93	76	80	70	83	80	82	1
19	13.1.03.02.0020	Mohammad Saiful Huda	60	75	79	70	76	80	78	4
20	13.1.03.02.0178	Muhamad Febrian Y.	88	79	84	70	80	82	80	5

Table 2 is the data testing on the classification of category selection of student thesis title. The attributes used are the value of information technology infrastructure planning course (PITI), the value of the software engineering course (RPL), the value of the network course, the data mining course grade, the value of the image processing course, the programming algorithm course (ALPRO), the value of the database course, and student interest.

3. Calculating Distance

Calculating the distance between attributes in the data testing with the distance attribute in data training using the formula below:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Where x1 = sample data (data training), x2 = test data (data testing), i = variable data, dist = distance, p = data dimension. In the calculation of variable x1 using data testing in table 2, variable x2 using training data in table 1, variable i is the attribute used is the value of information technology infrastructure planning courses (PITI), the value of software engineering courses (RPL) College network, data mining course grades, image processing course grades, programming algorithm course values (ALPRO), database course grades, and student interest. Here is the calculation of the distance on the test data to one:

$$d_i = \sqrt{\sum_{i=1}^p ((piti_2 - piti_1) + (rpl_2 - rpl_1) + (jaringan_2 - jaringan_1) + (d.mining_2 - d.mining_1) + (p.citra_2 - p.citra_1) + (alpro_2 - alpro_1) + (basdat_2 - basdat_1) + (minat_2 - minat_1))^2}$$

$$d_i = \sqrt{\sum_{i=1}^p ((80- 80) + (78 - 76) + (82- 77) + (80 - 79) + (88 - 85) + (80 - 79) + (81 - 79) + (5 - 5))^2}$$

$d_i = 6,69$

The distance calculation results for all training data on the first data testing as shown in Figure 2.

NO	ID	NAMA	PITI	RPL	Jaringan	Data Mining	Pengolahan Citra	Alpro	Basis Data	Minat	KATEGORI	PITI-P.Citra	Alpro-Minat	JUMLAH	JARAK
1	12.1.03.02.0082	Adhen Bagus Patro U.	80	76	77	79	85	79	79	5	Citra	40.25	4.50	44.75	6.69
2	12.1.03.02.0203	Aditya Bayu Saputro	91	77	85	79	85	81	84	1	SI	137.30	29.25	166.55	12.91
3	12.1.03.02.0428	Ahmad Basyar Zakia	80	77	79	79	80	80	84	5	Citra	74.38	12.25	86.63	9.31
4	12.1.03.02.0208	Aldila Anggun Sasmita	70	80	79	79	80	81	84	3	Jaringan	174.06	16.50	190.56	13.80
5	12.1.03.02.0034	Andika Susanto	82	85	80	83	70	81	84	2	RPL	389.06	21.50	410.56	20.26
6	12.1.03.02.0035	Andri Eko Umarudin	62	80	77	85	60	80	81	4	Mining	1162.00	1.25	1163.25	34.11
7	12.1.03.02.0262	Angga Aditya Sardi	70	83	81	83	70	82	84	2	RPL	459.00	23.50	482.50	21.97
8	12.1.03.02.0062	Anggara Lexa Prabawa	79	88	80	83	75	81	81	1	RPL	280.84	17.00	297.84	17.26
9	12.1.03.02.0227	Anggie Wibowo	69	80	77	80	70	80	80	2	RPL	471.56	9.50	481.06	21.93
10	12.1.03.02.0182	Aplil Lia Tn Winarsih	91	78	80	81	85	81	86	1	SI	122.56	46.50	169.06	13.00
11	12.1.03.02.0429	Aprilina Pratiwi	88	82	79	75	85	80	78	1	SI	123.00	21.06	144.06	12.00
12	12.1.03.02.0367	Ari Nugroho	88	76	80	77	88	81	82	5	Citra	83.25	1.25	84.50	9.19
13	12.1.03.02.0073	Astniani Agus Setyowati	91	83	80	83	85	81	81	1	SI	155.50	17.00	172.50	13.13
14	12.1.03.02.0395	Aviv Danar Putra P.	78	80	80	75	70	81	81	3	Jaringan	358.03	4.31	362.34	19.04
15	12.1.03.02.0136	Azis Zakana	60	76	79	82	60	82	81	4	Mining	1200.50	5.06	1205.56	34.72
16	12.1.03.02.0321	Bayu Nur Rozikin	77	85	77	77	70	80	79	2	RPL	409.36	10.56	419.92	20.49
17	12.1.03.02.0210	Bayu Putra Prasetya	65	76	80	74	60	80	83	3	Jaringan	1051.25	9.31	1060.56	32.57
18	12.1.03.02.0122	Debit Bagus Dwicahya	68	83	80	73	70	80	82	2	RPL	546.00	10.00	556.00	23.58
19	12.1.03.02.0274	Delli Widianna Putri	74	81	79	73	80	82	81	5	Citra	165.03	2.31	167.34	12.94
20	12.1.03.02.0187	Deni Naviga Amadanto	79	80	84	80	75	82	80	4	Mining	180.61	6.00	186.61	13.66
>>DATA TESTING															
NO	ID	NAMA	PITI	RPL	Jaringan	Data Mining	Pengolahan Citra	Alpro	Database	Minat					
1	13.1.03.02.0347	Achmad Misbakhul Huda	80	78	82	80	88	80	81	5					>> data testing 1

Figure 2. Distance Calculation

4. Determine K=5

At this stage determined the amount of data to be taken as the closest distance.

5. Sorting the data from the smallest distance to the largest distance, then taken 5 data with the smallest distance / shortest (K = 5). Figure 3 shows the result of calculating the sorted distance.

NO	ID	NAMA	PITI	RPL	Jaringan	Data Mining	Pengolahan Citra	Alpro	Database	Minat	KATEGORI	PITI-P.Citra	Alpro-Minat	JUMLAH	JARAK
1	12.1.03.02.0082	Adhen Bagus Putro U.	80	76	77	79	85	79	79	5	Citra	40.25	4.50	44.75	6.69
2	12.1.03.02.0367	Ari Nugroho	88	76	80	77	88	81	82	5	Citra	83.25	1.25	84.50	9.19
3	12.1.03.02.0428	Ahmad Basyar Zakia	80	77	79	79	80	80	84	5	Citra	74.38	12.25	86.63	9.31
4	12.1.03.02.0429	Aprilina Pratiwi	88	82	79	75	85	80	78	1	SI	123.00	21.06	144.06	12.00
5	12.1.03.02.0203	Aditya Bayu Saputro	91	77	85	79	85	81	84	1	SI	137.30	29.25	166.55	12.91
6	12.1.03.02.0274	Delhi Widianna Putri	74	81	79	73	80	82	81	5	Citra	165.03	2.31	167.34	12.94
7	12.1.03.02.0182	April Lia Tri Winarsih	91	78	80	81	85	81	86	1	SI	122.56	46.50	169.06	13.00
8	12.1.03.02.0073	Astriani Agus Setyowati	91	83	80	83	85	81	81	1	SI	155.50	17.00	172.50	13.13
9	12.1.03.02.0187	Deni Naviga Amadanto	79	80	84	80	75	82	80	4	Mining	180.61	6.00	186.61	13.66
10	12.1.03.02.0208	Aidila Anggun Sasmata	70	80	79	79	80	81	84	3	Jaringan	174.06	16.50	190.56	13.80
11	12.1.03.02.0062	Anggara Lexa Prabawa	79	88	80	83	75	81	81	1	RPL	280.84	17.00	297.84	17.26
12	12.1.03.02.0395	Aviv Danar Putra P.	78	80	80	75	70	81	81	3	Jaringan	358.03	4.31	362.34	19.04
13	12.1.03.02.0034	Andika Susanto	82	85	80	83	70	81	84	2	RPL	389.06	21.50	410.56	20.26
14	12.1.03.02.0321	Bayu Nur Rozikin	77	85	77	77	70	80	79	2	RPL	409.36	10.56	419.92	20.49
15	12.1.03.02.0227	Anggie Wibowo	69	80	77	80	70	80	80	2	RPL	471.56	9.50	481.06	21.93
16	12.1.03.02.0262	Angga Aditya Sardi	70	83	81	83	70	82	84	2	RPL	459.00	23.50	482.50	21.97
17	12.1.03.02.0122	Debit Bagus Dwicahya	68	83	80	73	70	80	82	2	RPL	546.00	10.00	556.00	23.58
18	12.1.03.02.0210	Bayu Putra Prasetya	65	76	80	74	60	80	83	3	Jaringan	1051.25	9.31	1060.56	32.57
19	12.1.03.02.0035	Andri Eko Umarudin	62	80	77	85	60	80	81	4	Mining	1162.00	1.25	1163.25	34.11
20	12.1.03.02.0136	Azis Zakaria	60	76	79	82	60	82	81	4	Mining	1200.50	5.06	1205.56	34.72
NO	ID	NAMA	PITI	RPL	Jaringan	Data Mining	Pengolahan Citra	Alpro	Database	Minat	KATEGORI	PITI-P.Citra	Alpro-Minat	JUMLAH	JARAK

Figure 3. Distance Calculation Result

Based on figure 3 sequence of the first shortest distance to the fifth, there are three categories of images and two categories of SI, so the first data testing by the name of student Achmad Misbakhul Huda included in the category of Image (*citra*).

3.2. System Design

In the classification system, the selection of students undergraduate thesis title using knn method there are two users namely lecturers and employees of Informatics Engineering Study Program. Figure 4 below shows the design of a system diagram context.

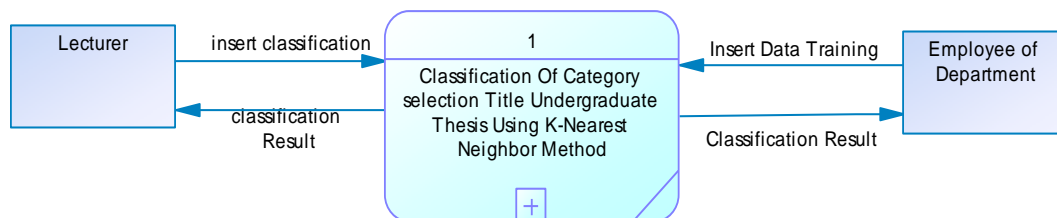


Figure 4. Context Diagram System

In Figure 4 shows entity lecturer can classify thesis title selection using knn method and lecturer entity can also receive classification information. Employee Prodi entity can enter training data as data reference process calculation method knn and employee of Prodi also can receive information of result of classification. Furthermore, a more detailed system diagram shown in Figure 5 data flow diagram (DFD) classification of student undergraduate thesis title selection using k-nearest neighbor method.

Figure 5 shows the flow of data flow diagram system where entity employees Prodi login process then login data will be verified in the data store login and if verified then the user employees Prodi can proceed to the process insert, update, delete data training. Further training data will be stored in data store training data. Likewise, the lecturer enters the login process first then the login data will be verified based on user data on the data store login, if the login data is verified then the lecturer user can proceed to the insert, update, delete data testing process, then the testing data will be stored in store testing data.

Training data will then be used as a reference to the process of calculating the knn method to classify the existing data testing, then the data classification of student thesis title

using knn method stored in the data store classification data and then displayed on the process of classification results to be viewed by lecturer entities and employee department entities.

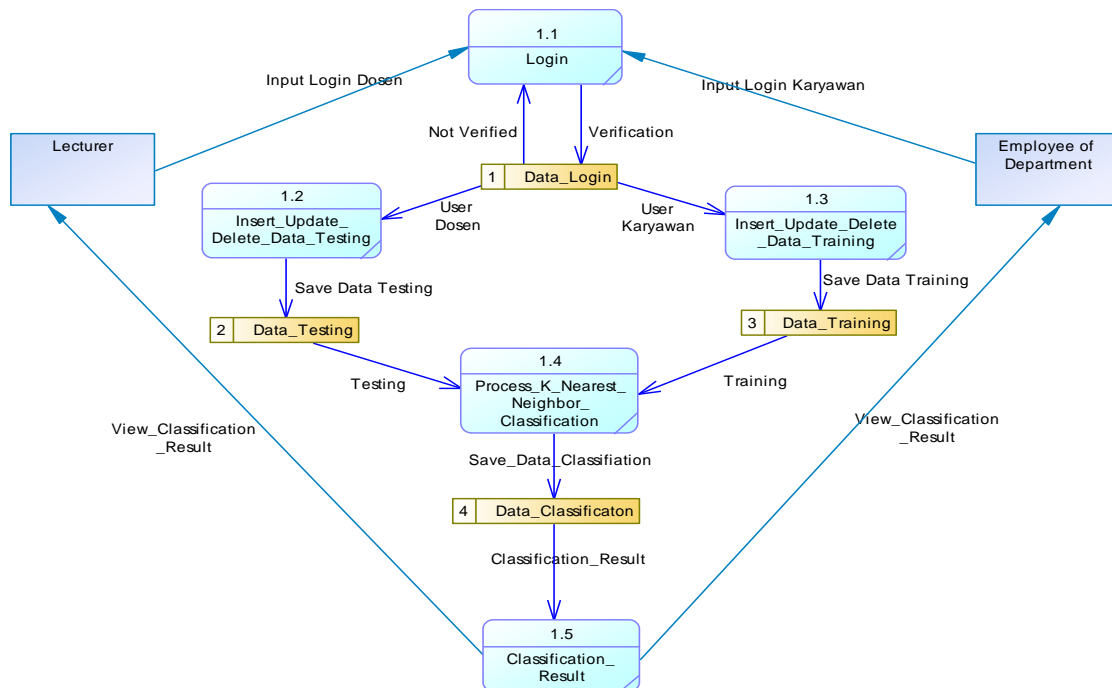


Figure 5. Data Flow Diagram System

The result of data store decomposition on classification system of student undergraduate thesis title using knn method described in detail on entity relationship diagram (ERD) in the Figure 6.

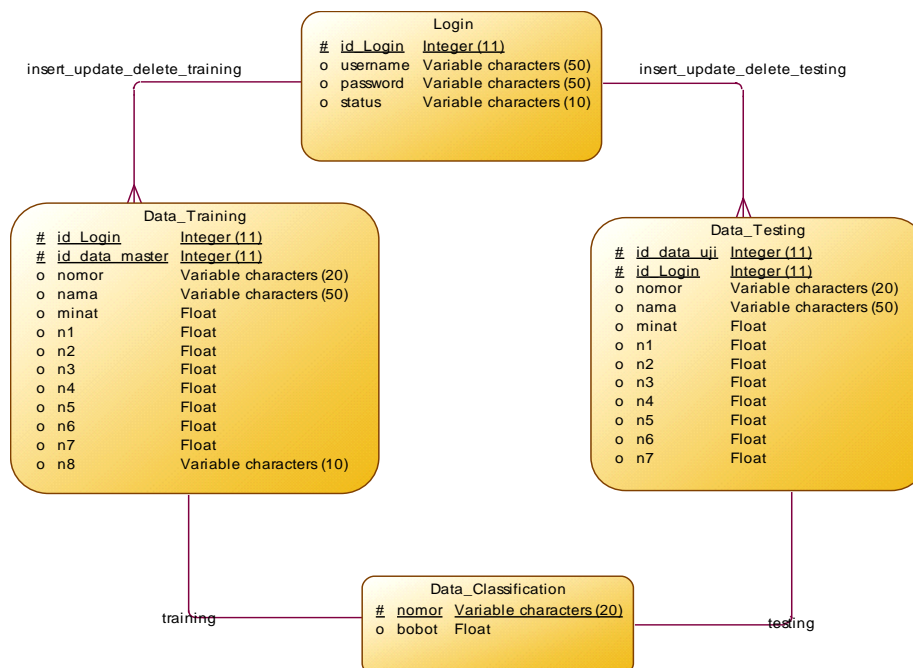


Figure 6. Entity Relationship Diagram

In entity relationship diagram Figure 6 there are four entities that are entity login, entity training data, data testing entity and entity data classification. The login entity has four attributes: the login id attribute as the primary key, the username attribute, the password attribute and the status attribute. The training data entity has twelve attributes: the id attribute of the master data as the primary key, the attribute of the number, the name attribute, the attribute of interest, the attribute n1 representing the value of the course of information technology infrastructure planning, the n2 attribute representing the value of the software engineering course, the attribute n3 Which represents the value of the network course, the n4 attribute representing the value of the data mining course, the n5 attribute representing the value of the image processing course, the n6 attribute representing the course value of the programming algorithm, the n7 attribute representing the value of the database course and the n8 attribute representing the undergraduate thesis category.

The data testing entity has eleven attributes namely the id attribute of the test data as the primary key, the attribute number, the name attribute, the attribute of interest, the attribute n1 representing the value of the course of information technology infrastructure planning, the n2 attribute representing the value of the software engineering course, the n3 attribute represents the value of the network course, the n4 attribute representing the value of the data mining course, the n5 attribute representing the value of the image processing course, the n6 attribute representing the course value of the programming algorithm and the n7 attribute representing the value of the database course. The data classification entity has two attributes namely the attribute number as the primary key and the weight attribute.

4. Conclusion

The conclusion from the making classification of category selection title undergraduate thesis using k-nearest neighbor method is as follows:

1. The research to apply k-nearest neighbor method that can classify the category of student undergraduate thesis title based on the interest and the values of the course of planning and infrastructure of information technology, software engineering, network, data mining, image processing, programming algorithm and database.
2. The results of the classification k-nearest neighbor method are grouped into categories of information systems, software engineering, networking, data mining and image processing.
3. The results of research used by lecturers and employees Prodi to facilitate the classification of the category of thesis in accordance with the field of interest and field of expertise of each student in Informatics Engineering Program Faculty of Engineering, University of Nusantara PGRI Kediri.

References

- [1] Government Regulation No.66 of 2010 on Amendment to Government Regulation No. 17 of 2010 on Management and Implementation of Education, 2010.
- [2] Big Indonesian Dictionary (KKBI), 2016, Online Version Dictionary, <http://kbbi.web.id/skripsi>,
- [3] Kusriani and Luthfi, E. T., *Data Mining Algorithm*, Andi Offset, Yogyakarta, 2009.
- [4] Ananda, K. R., *Decision Support System For Determining Thesis Head of Computer Information Technique Using Simple Additive Weighthing (SAW) Method*, Pelita Informatika Journal Budi Darma, 2014; Vol. VI No. 2 ISSN. 2301-9425.
- [5] Niswatin, R. K., *Decision Support System for New Student Placement Using K-Nearest Neighbor Method*, Cogito Smart Journal Manado, 2015; Vol. 1 No. 1 e-ISSN. 2477-8079.
- [6] Hastuti, K., *Comparative Analysis of Data Mining Classification Algorithm for Non-Active Student Prediction*, National Seminar on Information Technology and Applied Communication (Semantic), Semarang, June 2012; ISBN 979-26-0255-0.
- [7] Kusriani, *Concept and Application of Decision Support System*, Andi Offset, Yogyakarta, 2007.
- [8] Han, J. Dan Kamber, M., *Data Mining Concept and Techniques*, Morgan Kauffman, San Francisco, 2006.
- [9] Gorunescu, F., *Data Mining: Concepts, Models, and Techniques*, Springer, Verlag Berlin Heidelberg, 2011.
- [10] Larose, Daniel T., *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley and Sons. Inc, 2005.
- [11] Agusta, Yudi, *K-Means Application of Problems and Related Methods*, Journal of Systems and Informatics, 2007; Vol. 3 p. 47-60.