



## Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Mapreduce

Muhammad Awaluddin<sup>1</sup>, Muhammad Niswar<sup>2</sup>, Amil Ahmad Ilham<sup>3</sup>

<sup>1</sup>Sistem Komputer, STMIK Handayani Makassar

<sup>2,3</sup>Teknik Elektro, Universitas Hasanuddin

<sup>1</sup>m.awal@myself.com, <sup>2</sup>niswar@unhas.ac.id, <sup>3</sup>amil@unhas.ac.id

### Abstrak

Tujuan dalam penelitian ini adalah untuk membantu Perguruan Tinggi dalam peningkatan kualitas akademik dan membantu dalam menentukan kuota penerimaan mahasiswa baru dengan melakukan prediksi terhadap tingkat kelulusan dan predikat lulus mahasiswa. Metode yang digunakan untuk memprediksi predikat dan tingkat kelulusan mahasiswa pada penelitian ini adalah teknik Prediksi Data Mining dengan menggunakan MapReduce yang dimulai dengan Data selection atau pengumpulan data-data yang terdapat pada Pangkalan Data Pendidikan Tinggi (PDPT) dengan memanfaatkan web service. Pre-processing sebelum proses data mining dilaksanakan perlu dilakukan proses cleaning. Dataset akan diseleksi dari data yang mengandung missing value dan data dengan nilai yang redundan. Data Mining merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pada penelitian ini penerapan data mining menggunakan Hadoop-MapReduce. Informasi yang dihasilkan akan menampilkan dan memberikan informasi kemungkinan terhadap predikat dan tingkat kelulusan mahasiswa, dari sinilah pihak akademik dapat mengambil suatu keputusan terkait kebijakan Perguruan Tinggi. Dari penelitian yang telah dilakukan, maka didapat beberapa kesimpulan yaitu pengolahan data yang sangat cepat oleh framework Hadoop, data eksekusi paling lama adalah 2375 milidetik dengan jumlah total berkas 3,87 GB untuk mengeksekusi proses hitung IPK mahasiswa. Sedangkan Proses eksekusi tercepat adalah 105 milidetik untuk memproses hitung jumlah sks mahasiswa dengan total berkas 231 MB. Pada hasil pengujian untuk lulusan tahun 2015 diprediksi ada 1075 lulusan sedangkan mahasiswa lulus sebenarnya adalah 1098 mahasiswa sehingga didapatkan tingkat kesalahan prediksi ini adalah 2,09%.

Kata Kunci : Data Mining, Hadoop, MapReduce, Prediksi Kelulusan Mahasiswa, Web Service PDPT.

### Abstract

*The objectives of this research were to help the university in improving the academic quality and to help in determining the quota of new students acceptance by doing the prediction toward the graduation level and students' predicate. The method used to predict the predicate and the students graduation in this research was a data mining prediction technique by using Mapreduce which begun with data selection or data collection in the high education database through web service. Pre-processing before the process of data mining was done, the cleaning process was done. Dataset would be selected from the data contained the missing value and the data with redundant value. Data Mining was the stage to find the pattern or information in a group of data by using the technique and certain algorithm. The application of data mining in this research used Hadoop-Mapreduce. The produced information would display and gave information of the possibility toward the predicate and the students' graduation level, the academic side made a decision concerning with university wisdom. From the research has been done, it was concluded that the fastest data collection by Hadoop framework, the execution data at least was 2375 miles second with the total number of*



files about 3.87 GB to execute the calculation process of cumulative achievement index of students. Meanwhile, to the fastest execution process was 105 miles second to process the number of semester credit units of students with the total number of files was 231 MB. The result of testing for the graduation in 2015 was predicted about 1075 graduates, while the graduated students about 1098 students so that the error of prediction was 2.09%.

*Key Word : Data Mining, Hadoop, Mapreduce, the Prediction of Students Graduation, PDPT Web Service*

## 1. Pendahuluan

Pada era teknologi saat ini, data menjadi sesuatu yang penting dalam menjalankan berbagai hal, beberapa diantaranya; mengetahui tren pasar, mengetahui keinginan konsumen saat ini, meningkatkan hasil penjualan, dll. Hasil perubahan ini sangatlah besar, data pun diolah dengan lebih terkomputerisasi sehingga penyimpanan data dapat menghemat tempat dalam kantor dengan cara penyimpanan softcopy. Data yang tersimpan ini lama kelamaan menjadi sangat banyak dan besar sehingga semakin susah untuk digunakan, hal tersebut kemudian disebut “big data”.

Pemanfaatan data dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Pengolahan data dalam jumlah yang besar seringkali menyebabkan kegagalan dalam proses pengolahan data, hal ini menjadi masalah baru yang harus diselesaikan, teknologi Big Data kemudian hadir menjadi solusi pengolahan data dalam jumlah yang besar. Penggunaan metode data mining dengan menggunakan Hadoop-MapReduce diharapkan dapat memprediksi predikat dan tingkat kelulusan mahasiswa agar dapat membantu pihak akademik dalam peningkatan kualitas pendidikan dan juga membantu dalam penetapan kuota penerimaan mahasiswa baru.

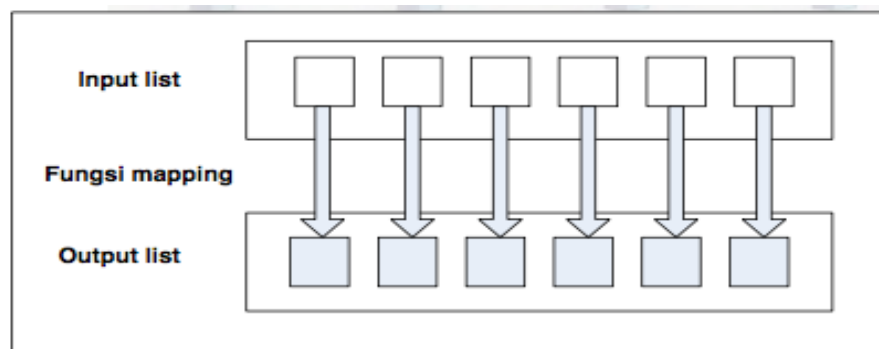
## 2. Metode Penelitian

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining, sering juga disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

Hadoop MapReduce adalah model pemrograman dan kerangka kerja perangkat lunak untuk menulis aplikasi yang cepat dalam memproses data dengan jumlah besar secara paralel pada kelompok besar dari node komputer. MapReduce menggunakan HDFS untuk mengakses segmen file dan untuk menyimpan hasil reduce. Hadoop Distributed File System (HDFS) adalah sistem penyimpanan utama yang digunakan oleh aplikasi Hadoop. HDFS adalah, seperti namanya, sistem file terdistribusi yang menyediakan akses throughput yang tinggi untuk data aplikasi dengan membuat beberapa replika blok data dan mendistribusikan mereka pada node komputasi seluruh cluster yang memungkinkan perhitungan handal dan cepat.

### 2.1 Map

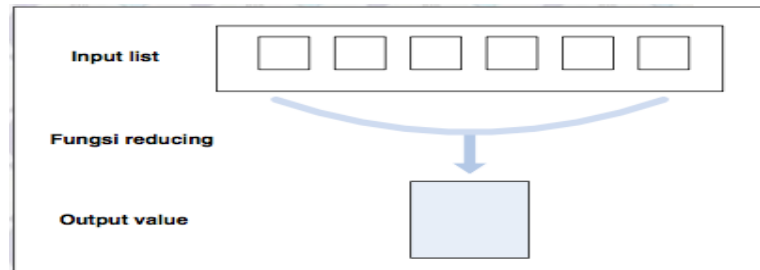
Map melakukan transformasi setiap data elemen input menjadi data elemen output. Map dapat dicontohkan dengan suatu fungsi toUpper(str) yang akan mengubah setiap huruf kecil (lowercase) menjadi huruf besar (uppercase). Setiap data elemen huruf kecil (lowercase) yang menjadi input dari fungsi ini akan ditransformasi menjadi data output elemen yang berupa huruf besar (uppercase). Map memiliki fungsi yang dipanggil untuk setiap input yang menghasilkan output pasanganintermediate <key, value>.



Gambar 1. Map menghasilkan output untuk setiap data elemen input

## 2.2 Reduce

Reduce adalah tahap yang dilakukan setelah mapping selesai. Reduce akan memeriksa semua value input dan mengelompokkannya menjadi satu value output. Reduce menghasilkan output pasangan intermediate. Sebelum memasuki tahap reduce, pasangan intermediate  $\langle \text{key}, \text{value} \rangle$  dikelompokkan berdasarkan key, tahap ini dinamakan tahap shuffle.



Gambar 2. Reduce mengelompokkan semua input list menjadi sebuah output value

## 3. Hasil dan Pembahasan

Dalam penelitian implementasi data mining untuk memprediksi predikat dan tingkat kelulusan mahasiswa dengan menggunakan Hadoop-MapReduce ini mengambil studi kasus pada Pangkalan Data Universitas Muhammadiyah Buton Kota Baubau.

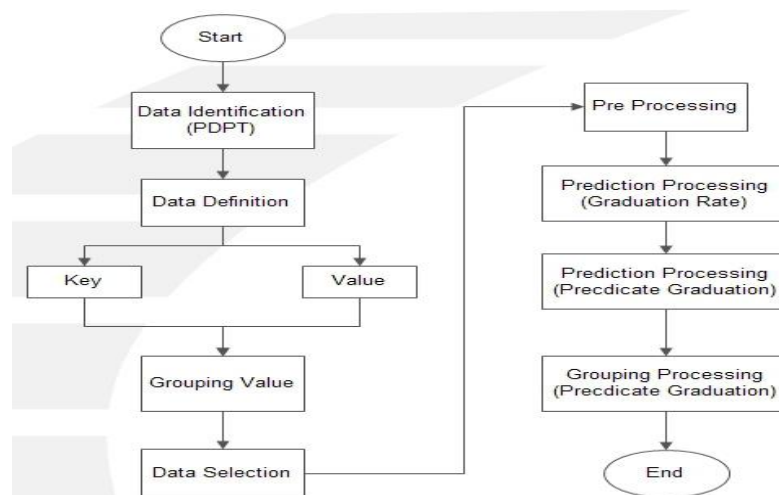
### 3.1. Metode Analisis Data

Metode yang digunakan untuk memprediksi predikat dan tingkat kelulusan mahasiswa pada penelitian ini adalah teknik Prediksi Data Mining dengan menggunakan MapReduce dengan tahapan :

- Data selection : Pada tahap ini penulis mengumpulkan data-data yang terdapat pada Pangkalan Data Pendidikan Tinggi (PDPT). Data yang diambil adalah data akademik mahasiswa dalam database PD-Dikti Feeder
- Pre-processing : Sebelum proses data mining dilaksanakan perlu dilakukan proses cleaning. Dataset akan diseleksi dari data yang mengandung missing value dan data dengan nilai yang redundan.
- Data Mining : Merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pada penelitian ini penerapan data mining menggunakan Hadoop-MapReduce
- Interpretation / Evaluation : Merupakan pola informasi yang dihasilkan dari proses data mining. Informasi yang dihasilkan akan menampilkan dan memberikan informasi kemungkinan terhadap predikat dan tingkat kelulusan mahasiswa, dari sinilah pihak akademik dapat mengambil suatu keputusan terkait kebijakan Perguruan Tinggi.

### 3.2. Perancangan Sistem

Adapun rancangan sistem pada penelitian implementasi data mining untuk memprediksi predikat dan tingkat kelulusan mahasiswa dengan menggunakan Hadoop-MapReduce ini di uraikan dalam Blok Diagram berikut.



Gambar 3. Rancangan Sistem



#### 4.2. Menghitung IPK

Indeks Prestasi Kumulatif (IPK) adalah indeks yang dihitung pada suatu program pendidikan lengkap atau pada akhir semester kedua dan seterusnya untuk seluruh mata kuliah yang diambilnya, yang dinyatakan dalam rentangan angka 0,00-4,00. Pada tahap ini dilakukan perhitungan Index Prestasi dari setiap semester yang telah diselesaikan ditambah dengan nilai prediksi semester akhir yang nilainya ditentukan dari nilai tengah (Median) dari keseluruhan IP Semester. Dengan persamaan sebagai berikut :

$$Me_1 = x\left(\frac{n+1}{2}\right) \quad \text{atau} \quad Me_2 = \frac{1}{2}\left(x\left(\frac{n}{2}\right) + x\left(\frac{n}{2}+1\right)\right)$$

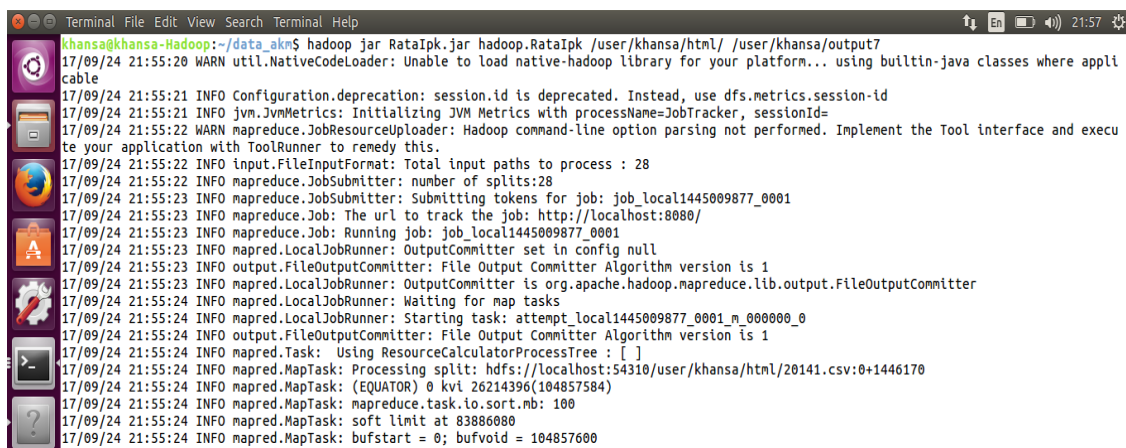
Keterangan:

$Me_1$  = Median dengan jumlah data ganjil

$Me_2$  = Median dengan jumlah data genap

$n$  = jumlah data

$x$  = nilai data yang telah disortir dari nilai terkecil



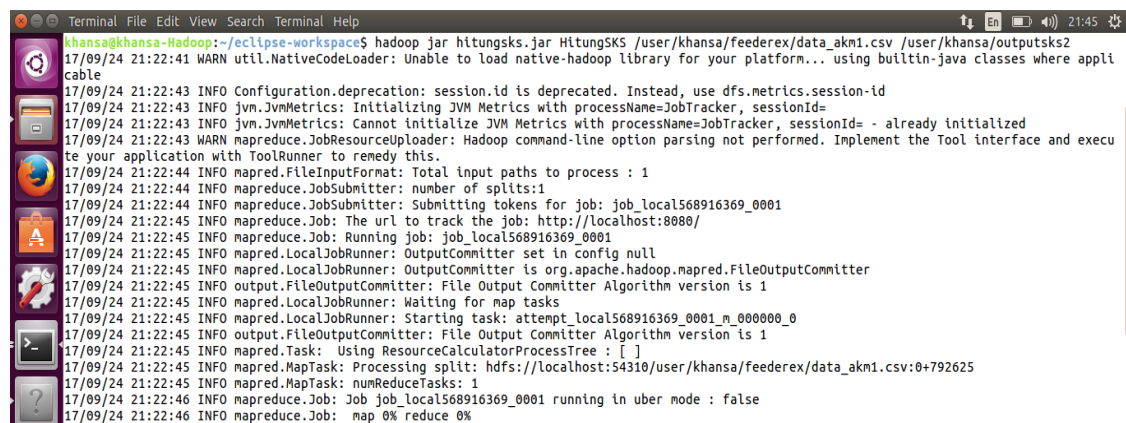
```
khansa@khansa-Hadoop:~/data_aks$ hadoop jar RataIpk.jar hadoop.RataIpk /user/khansa/html/ /user/khansa/output7
17/09/24 21:55:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/09/24 21:55:21 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/09/24 21:55:21 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/09/24 21:55:22 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/09/24 21:55:22 INFO input.FileInputFormat: Total input paths to process : 28
17/09/24 21:55:22 INFO mapreduce.JobSubmitter: number of splits:28
17/09/24 21:55:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1445009877_0001
17/09/24 21:55:23 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/09/24 21:55:23 INFO mapreduce.Job: Running job: job_local1445009877_0001
17/09/24 21:55:23 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/09/24 21:55:23 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/24 21:55:23 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/09/24 21:55:24 INFO mapred.LocalJobRunner: Waiting for map tasks
17/09/24 21:55:24 INFO mapred.LocalJobRunner: Starting task: attempt_local1445009877_0001_m_000000_0
17/09/24 21:55:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/24 21:55:24 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
17/09/24 21:55:24 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/user/khansa/html/20141.csv:0+1446170
17/09/24 21:55:24 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/09/24 21:55:24 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/09/24 21:55:24 INFO mapred.MapTask: soft limit at 83886080
17/09/24 21:55:24 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
```

Gambar 6. Proses Eksekusi Hitung IPK

Dari proses eksekusi Map-Reduce Framework di atas informasi yang didapat adalah data yang di proses pada tahapan Mapping adalah 89903 records, pada tahap Reduce 13515 records, dalam waktu yang sangat singkat yaitu 2375 milidetik dengan jumlah total berkas 4152037376 bytes atau setara dengan 3,87 Gigabytes.

#### 4.3. Menghitung Total SKS Mahasiswa

SKS adalah singkatan dari Satuan Kredit Semester. SKS sendiri adalah satuan “bobot” studi di setiap mata kuliah. Setiap mata kuliah memiliki jumlah SKS yang berbeda-beda. Satu mata kuliah biasanya “berbobot” 2-4 SKS. Pada tahapan ini dilakukan perhitungan jumlah sks (satuan kredit semester) mahasiswa berdasarkan sks per semester yang telah diselesaikan.



```
khansa@khansa-Hadoop:~/eclipse-workspace$ hadoop jar hitungsk.jar HitungSKS /user/khansa/feederex/data_aks1.csv /user/khansa/outputsk
17/09/24 21:22:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/09/24 21:22:43 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/09/24 21:22:43 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/09/24 21:22:43 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
17/09/24 21:22:43 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/09/24 21:22:44 INFO mapred.FileInputFormat: Total input paths to process : 1
17/09/24 21:22:44 INFO mapreduce.JobSubmitter: number of splits:1
17/09/24 21:22:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local568916369_0001
17/09/24 21:22:45 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/09/24 21:22:45 INFO mapreduce.Job: Running job: job_local568916369_0001
17/09/24 21:22:45 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/09/24 21:22:45 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
17/09/24 21:22:45 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/24 21:22:45 INFO mapred.LocalJobRunner: Waiting for map tasks
17/09/24 21:22:45 INFO mapred.LocalJobRunner: Starting task: attempt_local568916369_0001_m_000000_0
17/09/24 21:22:45 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/24 21:22:45 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
17/09/24 21:22:45 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/user/khansa/feederex/data_aks1.csv:0+792625
17/09/24 21:22:45 INFO mapred.MapTask: numReduceTasks: 1
17/09/24 21:22:46 INFO mapreduce.Job: Job job_local568916369_0001 running in uber mode : false
17/09/24 21:22:46 INFO mapreduce.Job: map 0% reduce 0%
```

Gambar 7. Proses Eksekusi Script Hitung SKS

Dari proses eksekusi Map-Reduce Framework di atas informasi yang didapat adalah data yang di proses pada tahapan Mapping adalah 4782 records, pada tahap Reduce 1211 records, dalam waktu yang sangat singkat yaitu 89 milidetik dengan jumlah total berkas 242360320 bytes atau setara dengan 231 Megabytes.

#### 4.4. Penentuan Predikat Mahasiswa

Sesuai Panduan Akademik Mahasiswa Universitas Muhammadiyah Buton yang merujuk pada Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015 Tentang Standar Nasional Pendidikan Tinggi yang di dalamnya mengatur tentang pemberian predikat mahasiswa adalah sebagai berikut :

Tabel 1. Index Kelulusan

Indeks Prestasi	Indeks Kelulusan
3,51 - 4,00	Pujian
3,01 - 3,50	Sangat Memuaskan
2,75 - 3,00	Memuaskan

Dari rangkaian proses di atas didapat hasil seperti dibawah ini :

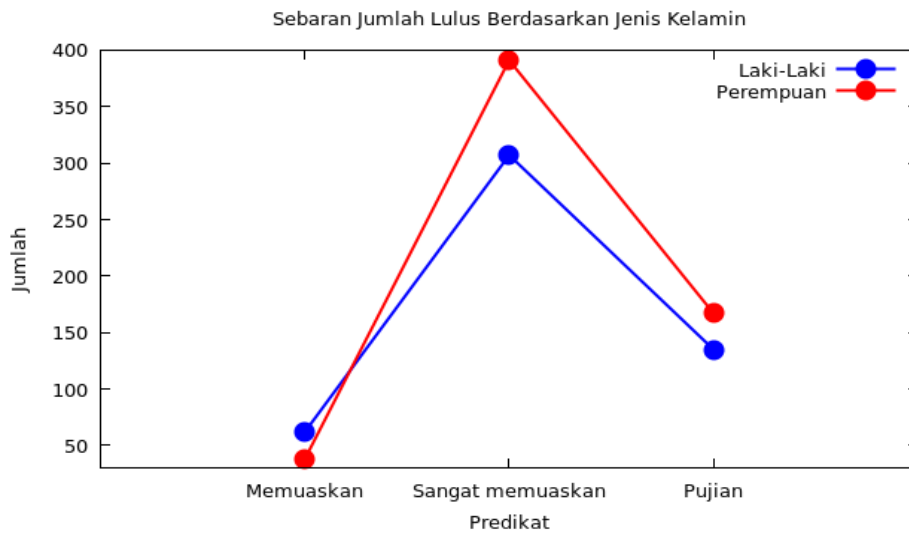
101201059	166	10	20121	2,75	Memuaskan	P
101201060	169	10	20121	2,75	Memuaskan	L
101201061	166	10	20121	2,76	Memuaskan	L
101201062	166	10	20121	2,76	Memuaskan	P
101201063	178	10	20121	2,76	Memuaskan	L
101201064	172	10	20121	2,76	Memuaskan	P
101201065	166	10	20121	2,78	Memuaskan	L
101201067	166	10	20121	2,78	Memuaskan	P
101201068	166	10	20121	2,78	Memuaskan	L
101201069	166	10	20121	2,79	Memuaskan	L
101201072	160	10	20121	2,79	Memuaskan	P
101201073	166	10	20121	2,81	Memuaskan	L
101201075	166	10	20121	2,82	Memuaskan	P
101201076	165	9	20121	2,82	Memuaskan	P
101201078	169	10	20121	2,82	Memuaskan	L
101201079	166	10	20121	2,83	Memuaskan	L

Gambar 8. Data hasil proses

Dapat dipaparkan sebaran Predikat mahasiswa lulus dalam bentuk tabel dan ditampilkan sebagai grafik dengan menggunakan GNU PLOT berikut ini :

Tabel 2. Sebaran Predikat Lulus Berdasarkan Jenis Kelamin

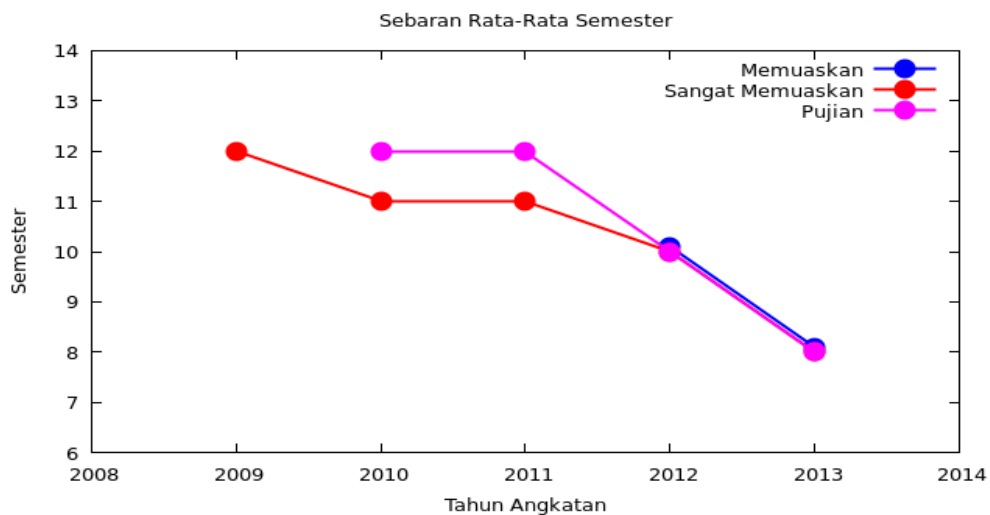
Predikat	Jenis Kelamin	
	Laki-laki	Perempuan
Memuaskan	62	38
Sangat Memuaskan	307	391
Pujian	135	167



Gambar 9. Grafik Sebaran Predikat Lulus Berdasarkan Jenis Kelamin

Tabel 3. Predikat Lulus Berdasarkan Rata-Rata Semester

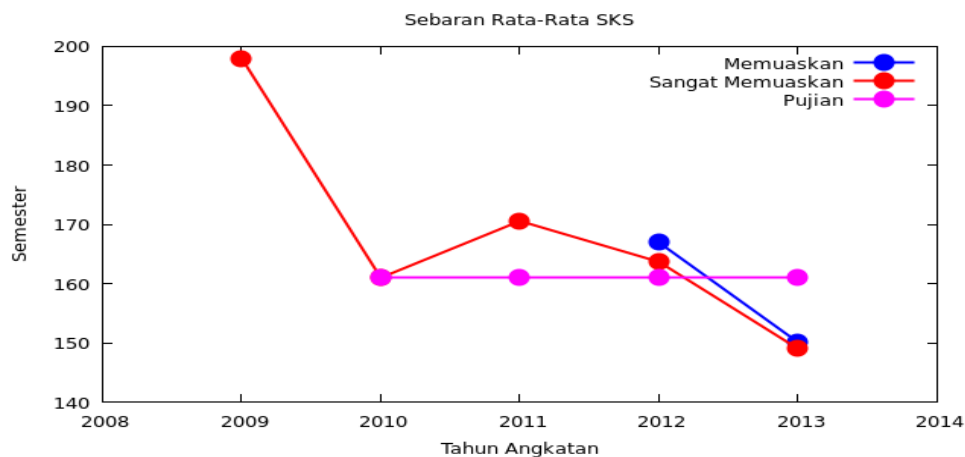
Angkatan	Predikat		
	Mumuaskan	Sangat Memuaskan	Pujian
2009	-	12	-
2010	-	11	12
2011	-	11	12
2012	10	10	10
2013	8	8	8



Gambar 10. Grafik Sebaran Predikat Lulus Berdasarkan Rata-Rata Semester

Tabel 4. Predikat Lulus Berdasarkan Rata-Rata Total SKS Lulus

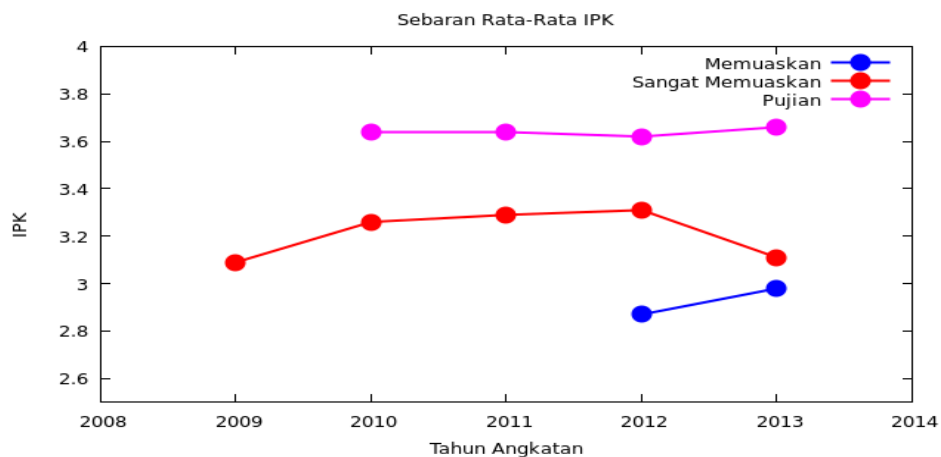
Angkatan	Predikat		
	Mumuaskan	Sangat Memuaskan	Pujian
2009	-	198	-
2010	-	161	161
2011	-	171	161
2012	167	164	161
2013	150	149	161



Gambar 11. Grafik Sebaran Predikat Lulus Berdasarkan Rata-Rata Total SKS Lulus

Tabel 5. Predikat Lulus Berdasarkan Rata-Rata IPK

Angkatan	Predikat		
	Memuaskan	Sangat Memuaskan	Pujian
2009	-	3,09	-
2010	-	3,26	3,64
2011	-	3,29	3,64
2012	2,87	3,31	3,62
2013	2,98	3,11	3,66



Gambar 12. Grafik Sebaran Predikat Lulus Berdasarkan Rata-Rata IPK

### 4.3 Pengujian

Pengujian dilakukan dengan melakukan prediksi pada lulusan tahun 2015 dengan jumlah lulusan sebenarnya 1098 mahasiswa. Didapatkan hasil prediksi lulusan sejumlah 1075 mahasiswa, lulusan sebenarnya adalah 1098 mahasiswa,  $1075 - 1098 = -23$  dari data tersebut terdapat selisih 23 mahasiswa. Sehingga perhitungan tingkat kesalahannya adalah  $23 \div 1098 = 0,0209$  atau 2,09%.

### 5. Kesimpulan

Pengolahan data yang sangat cepat oleh framework Hadoop, data eksekusi paling lama adalah 2375 milidetik dengan jumlah total berkas 3,87 GB untuk mengeksekusi proses hitung IPK mahasiswa. Sedangkan Proses eksekusi tercepat adalah 105 milidetik untuk memproses hitung jumlah sks mahasiswa dengan total berkas 231 MB. Pada hasil pengujian untuk lulusan tahun 2015 diprediksi ada 1075 lulusan sedangkan mahasiswa lulus sebenarnya adalah 1098 mahasiswa sehingga didapatkan tingkat kesalahan prediksi ini adalah 2,09%.



**Referensi :**

- [1] Lillyan Hadjaratie. Prediksi Dan Pemetaan Data Mahasiswa Fakultas Teknik Universitas Negeri Gorontalo Menggunakan Pendekatan Data Mining. Jurusan Teknik Informatika Universitas Negeri Gorontalo. 2014.
- [2] Siti Nurhayati, Kusri, Emha Taufiq Luthfi. Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine. Universitas Yapis Papua, Magister Teknik Informatika STMIK AMIKOM Yogyakarta. 2015.
- [3] Xuan Liu, Xiaoguang Wang, Stan Matwin, Nathalie Japkowicz. Meta-MapReduce for scalable data mining. Journal of Big Data. Faculty of Computer Science, Dalhousie University. Canada. 2012.
- [4] Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. Google, Inc. 2004.
- [5] Asniar Tj. Penggunaan Big Data Analytic Di Perguruan Tinggi. Fakultas Ilmu Terapan, Universitas Telkom. Bandung. 2015.
- [6] Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015 Tentang Standar Nasional Pendidikan Tinggi

