

e-ISSN : 2722-4058

JIFORTY

Journal of Information and Information Security



Vol.2 No. 1
Edisi Juni 2021

Program Studi Informatika
Universitas Bhayangkara Jakarta Raya
Web : <http://ejurnal.ubharajaya.ac.id/index.php/jiforty>
e-mail : jiforty.tif@ubharajaya.ac.id



[Beranda](#) / Dewan Editor

Dewan Editor

EDITOR IN CHIEF

Dwi Budi Srisulistiowati, Kom., M.M. (Google Scholar ID: [3sXU6CMAAAAJ](#), Universitas Bhayangkara Jakarta Raya, Indonesia)

MANAGING EDITOR

Fata Nidaul Khasanah, S.Kom., M.Eng (Scopus ID: [57189353040](#), Google Scholar ID: [H_Jkce8AAAAJ](#), Universitas Bhayangkara Jakarta Raya, Indonesia)

EDITORIAL BOARD MEMBERS

Kusdarnowo Hartanto, S.Kom., M.Kom. (Google Scholar ID: xxxxxxx, Universitas Bhayangkara Jakarta Raya, Indonesia)

Acmad Noe'man, S.Kom, M.Kom. (Google Scholar ID: xxxxxxx, Universitas Bhayangkara Jakarta Raya, Indonesia)

Siti Setiawati, S.Pd , M.Pd. (Google Scholar ID: xxxxxxx, Universitas Bhayangkara Jakarta Raya, Indonesia)

Lukmanul Hakim, S.Kom., M.Kom. (Google Scholar ID: xxxxxx, STMIK Antar Bangsa, Banten, Indonesia)

Jaja, S.Kom., M.Kom. (Google Scholar ID: xxxx, Universitas Subang, Subang, Indonesia)

Dian Ade Kurnia, M.Kom. (Google Scholar ID: [LPjnkocAAAAJ](#), STIMIK IKMI Cirebon, Cirebon, Indonesia)

TECHNICAL EDITOR

Wowon Priatna, S.T., M.T. (Google Scholar ID: [prtZKwIAAAAJ](#), Universitas Bhayangkara Jakarta Raya, Indonesia)





Journal of Informatics and Information Security (JIFORTY)

E-ISSN: 2722-4058

Sekretariat Redaksi Journal of Informatics and Information Security (JIFORTY)

e-mail:jiforty.tif@ubharajaya.ac.id

Jl. Perjuangan No.81, Marga Mulya, Kec. Bekasi Utara, Kota Bks, Jawa Barat 17143





Beranda / Arsip / Vol 2 No 1 (2021): Juni 2021

Vol 2 No 1 (2021): Juni 2021

DOI: <https://doi.org/10.31599/jiforty.v2i1>

Diterbitkan: 2021-06-30

Artikel

Perancangan Aplikasi Pengendalian Ruang Kelas Berbasis Mobile

Dian Hartanti, Bayu Pratama, Dwipa Handayani

pdf (English)

Model Kepuasan Orang Tua/Wali Siswa Berdasarkan Peningkatan Kualitas Pelayanan Sekolah Secara Berkelanjutan (Studi Kasus TK. Bina Mulia)

Muhamad Khaerudin, Dwi Budi Srisulistiowati, Siti Setiawati

pdf (English)

Analisis Biaya Standar Sebagai Alat Pengendalian Biaya Tenaga Kerja Langsung pada PT Masaiki Jakarta

Rani Suryani, Dwi Budi Srisulistiowati

pdf (English)

Aplikasi Kehadiran Karyawan Berbasis Android Menggunakan QR Code Scanning dan Location Based Service

Henny Leidiyana, Iskandar Yusuf

pdf (English)

Perancangan Sistem Informasi Persediaan Barang Menggunakan Visual Basic 2010 (Vb.Net) Pada PT. Solusi Rekatama Makmur

 Sri Rejeki, Ismasari Nawangsih, Siti Setiawati

 pdf (English)

Diagnosa COVID-19 Chest X-Ray Menggunakan Arsitektur Inception Resnet

 Adhitho Satyo Bayangkari Karno Satyo, Dodi Arif, Indra Sari Kusuma Wardhana, Eka Sally Moreta

 pdf (English)

Pemberdayaan Model Evaluasi CSE-UCLA Pada Program Sertifikasi Kompetensi BNSP Bidang Network Administrator Madya Dan Pemrograman Basis Data Di LSP-P1 Universitas Bhayangkara Jakarta Raya

 Ahmad Fathurrozi, R. Wisnu Prio Pamungkas

 pdf (English)

Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning

 Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra

 pdf (English)

Penanganan Risiko Keamanan Informasi Aplikasi Webmarket Berdasarkan ISO 27001:2013 (Study Kasus pada PT Sanjaya Citra Anugrah)

 M. Hadi Prayitno

 pdf (English)

Pembangunan Aplikasi E-Magang Perguruan Tinggi dengan Memanfaatkan SDLC SCRUM pada Agile Project Management

 Prio Kustanto, R. Wisnu P. Pamungkas, Ahmad Fathurrozi

 pdf (English)

Penerapan Soft Sistem Methodology Pada Pengembangan Sistem Informasi Akademik (Studi Kasus di SMKN 6 Kota Bekasi)

 Ani Yoraeni Hasan, Henny Leidiyana

 pdf (English)

Sistem Pengelolaan Kegiatan Perusahaan Menggunakan Identifikasi QR Code

👤 Dani Yusuf, Joni Warta, Sri Rejeki

📄 pdf (English)



Journal of Informatics and Information Security (JIFORTY)

E-ISSN: 2722-4058

Sekretariat Redaksi Journal of Informatics and Information Security (JIFORTY)

e-mail:jiforty.tif@ubharajaya.ac.id

Jl. Perjuangan No.81, Marga Mulya, Kec. Bekasi Utara, Kota Bks, Jawa Barat 17143



Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning

Rakhmat Purnomo^{1,*}, Wowon Priatna¹, Tri Dharma Putra¹

¹ Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Perjuangan 081, Marga Mulya, Bekasi Utara; 02188955882; e-mail: rakhmat.purnomo@dsn.ubharajaya.ac.id, wowon.priatna@dsn.ubharajaya.ac.id, tridharma.putra@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: rakhmat.purnomo@dsn.ubharajaya.ac.id

Diterima: 10 Juni 2021; Review: 26 Juni 2021; Disetujui: 29 Juni 2021; Diterbitkan: 3 Juli 2021

Abstract

The dynamics of higher education are changing and emphasize the need to adapt quickly. Higher education is under the supervision of accreditation agencies, governments and other stakeholders to seek new ways to improve and monitor student success and other institutional policies. Many agencies fail to make efficient use of the large amounts of available data. With the use of big data analytics in higher education, it can be obtained more insight into students, academics, and the process in higher education so that it supports predictive analysis and improves decision making. The purpose of this research is to implement big data analytical to increase the decision making of the competent party. This research begins with the identification of process data based on analytical learning, academic and process in the campus environment. The data used in this study is a public dataset from UCI machine learning, from the 33 available variables, 4 variables are used to measure student performance. Big data analysis in this study uses spark apace as a library to operate pyspark so that python can process big data analysis. The data already in the master slave is grouped using k-mean clustering to get the best performing student group. The results of this study succeeded in grouping students into 5 clusters, cluster 1 including the best student performance and cluster 5 including the lowest student performance.

Keywords: Machine Learning, Big Data Analytical, Learning Academic, k-mean Clustering Python, Hadoop, apache spark.

Abstrak

Dinamika pendidikan tinggi sedang berubah dan menekankan kebutuhan untuk beradaptasi dengan cepat. Pendidikan tinggi berada di bawah pengawasan lembaga akreditasi, pemerintah, dan pemegang kepentingan lainnya untuk mencari cara baru untuk meningkatkan dan memantau keberhasilan mahasiswa dan kebijakan kelembagaan lainnya. Banyak lembaga gagal memanfaatkan sejumlah besar data yang tersedia secara efisien. Dengan penggunaan big data analytic di perguruan tinggi maka dapat diperoleh wawasan yang lebih tentang mahasiswa, akademisi, dan proses di perguruan tinggi sehingga mendukung analisis prediksi dan peningkatan pengambilan keputusan. Tujuan penelitian ini adalah untuk implemetasi big data analytical untu menigkat pegambilan keputusan pihak kampu. Penelitian ini dimulai dengan identifikasi data-data proses berdasarkan learning analytical, academic dan proses di lingkungan kampus. Data yang digunakan dalam penelitian ini adalah dataset public dari bersumber dari UCI machine learning, dari 33 variabel tersedia digunakan 4 variabel untuk mengukur kinerja mahasiswa. Big data analisis dalam penelitian ini menggunakan spark apace sebagai library untuk mengoperasikan pyspark agar python dapat mengolah big data analisis. Data yang sudah ada di master slave dilakukan pengelompokan data menggunakan k-mean clustering untuk medapatkan kelompok mahasiswa yang kinerja terbaik. Hasil dari penelitian ini

berhasil mengelompokkan mahasiswa menjadi 5 cluster, cluster 1 termasuk dalam kinerja mahasiswa terbaik dan cluster 5 termasuk kinerja mahasiswa yang terendah.

Kata kunci: Machine Learning, Big Data Analytical, Learning Academic, k-mean Clustering Python, Hadoop, apache spark.

1. Pendahuluan

Perkembangan Teknologi digital di berbagai ruang dan waktu terus menghasilkan data dalam jumlah besar. Big Data menggambarkan pertumbuhan yang signifikan dalam volume dan variasi data yang tidak dapat lagi dikelola menggunakan database tradisional. Dengan bantuan analitik, jumlah data yang tampaknya berbeda dan heterogen ini dapat diproses untuk pola, yang pada dapat menghasilkan wawasan penting untuk pengambilan keputusan (Daniel, 2015) Dampak Digitalisasi data telah membuka peluang penggunaan big data di perguruan tinggi. Nilai big data terletak pada hasil analisis dan prediksi atau tindakan yang diambil dari hasil analisis dan prediksi tersebut (ADMIN SEVIMA, 2019). Dinamika pendidikan tinggi sedang berubah dan menekankan kebutuhan untuk beradaptasi dengan cepat. Pendidikan tinggi berada di bawah pengawasan lembaga akreditasi, pemerintah, dan pemegang kepentingan lainnya untuk mencari cara baru untuk meningkatkan dan memantau keberhasilan siswa dan kebijakan kelembagaan lainnya (Tulasi, 2013)

Pendidikan masa depan lebih sering dihubungkan dengan teknologi baru, pendidikan tinggi beroperasi dengan lingkungan semakin kompleks (Asniar, 2015) seperti perangkat komputasi di mana-mana, desain ruang kelas yang fleksibel. Karena institusi pendidikan tinggi mempunyai banyak data tentang mahasiswa, dan basis data catatan mahasiswa menjadi lebih kompleks untuk diakses, maka Perguruan Tinggi harus mendata semua data terkait akademik dari berbagai kegiatan seperti data mahasiswa, data registrasi, data asesmen dan lainnya. Untuk mempermudah pendidikan tinggi dalam mengakes yang dapat dihasilkan oleh mahasiswa adalah menerapkan analytical big data dalam e-learning analytical, dimana institusi dapat menentukan mahasiswa yang sukses atau mahasiswa yang drop out (Asniar, 2015).

Penelitian (Murumba & Micheni, 2017) menerapkan penerapan big data mengeksplorasi atribut big data yang relevan dengan institusi pendidikan. Penelitian (Tulasi, 2014) menerapkan big data analisis dalam analisa pembelajaran dengan menganalisa data statis dan dinamis dilingkungan belajar-mengajar untuk memungkinkan intervensi tepat waktu oleh pendidik sedangkan (Kumar Sinha & Singh, 2019) rekomendasi analisis big data dalam mencapai keberhasilan mahasiswa diberbagai perguruan tinggi. Penelitian (Daniel, 2015) menerapkan big data untuk rekomendasi project untuk proyek big data untuk tingkat kelembagaan institusi.

Beberapa Penelitian menggunakan kombinasi big data dan machine learning dalam menyelesaikan permasalahan data besar diantaranya: menggunakan machine learning dalam mengolah data besar untuk sisi keamanan (Kaur, Sharma, & Mittal, 2018) untuk memperkirakan

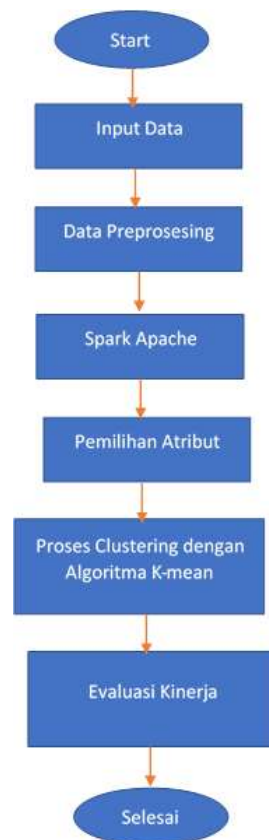
permintaan pariwisata kapal pesiar Tiongkok (Xie, Qian, & Wang, 2021) pengenalan wajah (Vinay et al., 2015) kerangka sistem untuk otonomi (Jamshidi, 2017)

Dari latar belakang dan penelitian tujuan penelitian ini adalah untuk implementasi big data analisis diperguruan tinggi dengan berfokus kepada learning academic dalam menentukan prediksi kinerja mahasiswa menggunakan k-mean clustering.

2. Metode Penelitian

Dalam Penelitian ini menggunakan dataset yang bersifat public dapat Students Academic Performance yang bersumber dari <https://archive.ics.uci.edu> jumlah 32 attribut diambil 4 yang akan digunakan dalam variable yang akan diolah oleh k-mean clustering.

Tahapan dalam penelitian ini dapat dilihat pada gambar 1.



Sumber : Hasil Penelitian (2021)

Gambar 1. Tahapan Penelitian

2.1. Input Data

Data yang telah diperoleh berupa format excel akan dirubah dan disesuaikan dengan format csv yang akan digunakan pada pemograman python, selanjutnya akan diinput kedalam directory folder yang nantinya akan ditampilkan pada directory Jupiter.

2.2. Data Preprocessing

Dalam tahapan persiapan pemrosesan data adalah mengecek data, memastikan tidak record, attribute yang kosong atau pun format sudah sesuai yang akan diolah python. Data yang akan diolah adalah data student performance sebagai bagian dari academic analytical dalam penerapan big data analytical dalam perguruan tinggi.

2.3. Spark Apache

Dalam tahapan ini adalah instalasi spark apache dan perangkat lunak lainnya. Spark apache ini berfungsi untuk mengoperasikan big data analisis. Dimana beberapa software pendukung yang harus diinstall adalah: Anaconda, Spark apache, Hadoop, Java development kit (JDK) dan skala. Dalam tahapan ini setelah semua perangkat lunak telah terinstall dan spark apache telah berhasil digunakan. Selanjutnya akan membuat server dalam spark apache yang dinamakan master slave. Master slave yang telah berhasil dikoneksikan selanjutnya IP address akan digunakan untuk membuat koneksi worker (client).

2.4. Data Selection (pemilihan Atribut)

Data set dengan format csv dapat diproses dengan Bahasa pemrograman Python. Maka, rentang data yang akan diolah harus didefinisikan terlebih dahulu. Melakukan seleksi rentang data adalah impor fungsi arrange dari library numpy serta membuat sebuah variabel yang menampung fungsi arrange. Definisikan secara numerik jumlah baris data yang ingin diseleksi. Proses ini akan menghasilkan variabel yang menampung data yang telah diseleksi.

2.5. K-mean Clustering

Untuk menjalankan big data analisis dengan Algoritma k-mean Clustering dipemrograman python dibutuhkan beberapa tahapan:

1. Mengimport pyspark (sebuah library big data menggunakan spark apache)
2. Mengimport library machine learning (MLlib) yang dibutuhkan oleh python untuk memproses machine learning
3. Mengimport library k-mean clustering
4. Menentukan centroid (titik pusat cluster) menggunakan mean masing-masing kelompok data
5. Mealokasikan data kepada centroid terdekat
6. Proses berulang Melakukan iterasi untuk memastikan tidak ada data yang pindah cluster.

2.6. Evaluasi Kinerja

Pada tahapan ini akan menguji kinerja dari hasil prediksi k-mean clustering apakah sudah efektif sehingga dapat digunakan sebagai rekomendasi model untuk digunakan.

3. Hasil dan Pembahasan

3.1. Perancangan Big Data

3.1.1. Instalasi Spark apache

Dalam penelitian ini untuk menjalankan Big Data digunakan spark apache untuk memproses data berskala besar. Spark apache menyediakan library untuk mengoperasikan python, distribusi skala, dan menampung Hadoop. Spark apache yang digunakan adalah versi 2.4.7 digunakan untuk membuat master Komputer sebagai server dan mengoperasikan akses client menggunakan worker. Berikut tampilan spark apache yang sudah terinstall di computer dapat dilihat pada gambar 2.



```

at org.apache.hadoop.util.StringUtils.<init>(StringUtils.java:88)
at org.apache.hadoop.security.SecurityUtil.getAuthenticationHeader(SecurityUtil.java:611)
at org.apache.hadoop.security.UserGroupInformation.initialize(UserGroupInformation.java:179)
at org.apache.hadoop.security.UserGroupInformation.ensureInitialized(UserGroupInformation.java:261)
at org.apache.hadoop.security.UserGroupInformation.getLoginUser(UserGroupInformation.java:771)
at org.apache.hadoop.security.UserGroupInformation.getCurrentUser(UserGroupInformation.java:781)
at org.apache.hadoop.security.UserGroupInformation.getCurrentUser(UserGroupInformation.java:834)
at org.apache.spark.util.Utils$anonfun$getCurrentUserName$.apply(Utils.scala:2422)
at org.apache.spark.util.Utils$anonfun$getCurrentUserName$.apply(Utils.scala:2422)
at scala.Option.getOrElse(Option.scala:121)
at org.apache.spark.util.Utils.getCurrentUserName(Utils.scala:2422)
at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
at org.apache.spark.deploy.SparkSubmit.<init>(SparkSubmit.scala:148)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$anonfun$1(SparkSubmit.scala:148)
at org.apache.spark.deploy.SparkSubmit$$anonfun$prepareSubmitEnvironment$.apply(SparkSubmit.scala:156)
at scala.Option.map(Option.scala:146)
at org.apache.spark.deploy.SparkSubmit.prepareSubmitEnvironment(SparkSubmit.scala:156)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$anonfun$main(SparkSubmit.scala:174)
at org.apache.spark.deploy.SparkSubmit.$anonfun$main$1(SparkSubmit.scala:161)
at org.apache.spark.deploy.SparkSubmit.$anonfun$main$2(SparkSubmit.scala:161)
at org.apache.spark.deploy.SparkSubmit$$anonfun$2(SparkSubmit.scala:166)
at org.apache.spark.deploy.SparkSubmit$$anonfun$2(SparkSubmit.scala:166)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:174)
at org.apache.spark.deploy.SparkSubmit.$anonfun$main$3(SparkSubmit.scala:174)
11/02/20 22:21:55 WARN NativeCodeLoader: unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context available as 'sc' (master = local[*]), app id = local-1442787261162).
Spark session available as 'spark'
Welcome to

Spark version 2.4.7

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_101)
Type in expressions to have them evaluated.
Type :help for more information.
  
```

Sumber : Hasil Penelitian (2021)

Gambar 2. Hasil instalasi Spark Apache

3.1.2. Instalasi apache Hadoop

Apache Hadoop merupakan salah satu distribusi big data yang dikembangkan oleh Apache Software Foundation. Apache Hadoop ini dibangun dengan lisensi free dan open source. Dalam penelitian ini digunakan Apache Hadoop sebagai framework untuk Big Data dengan versi 2.7.0. Instalasi dan konfigurasi dilakukan pertama kali pada komputer yang bertindak sebagai hadoop-master (single node), untuk menjadikan multimode cluster akan dilakukan sinkronisasi dengan komputer slave dan dengan melakukan perubahan konfigurasi yang minimal pada node. Secara garis besar tahapan instalasi dan konfigurasi Big Data Hadoop dengan Install Java Development Kit (JDK), Buat user dan group untuk Hadoop, Konfigurasi firewall dinonaktifkan dan Instalasi Hadoop dan konfigurasi Hadoop Environment pada node master dan slave.

3.1.3. Konfigurasi Environment

Pada tahapan ini adalah setting environment variable pada my computer agar big data dapat digunakan. Setting environment dapat dilihat pada tabel 1.

Tabel 1. Setting Environment

Variable	Fungsi
PYSPARK_DRIVER_PYTHON = E:\phyton\scripts\jupyter.exe	Untuk menjalankan python pada spark apache. Dimana editor menggunakan jupyter.exe
PYSPARK_DRIVER_PYTHON_OPTS = notebook	Tempat menulis code
PYSPARK_PYTHON = E:\phyton\python.exe	Agar pyspark dapat dijalankan
C:\Program Files (x86)\Common Files\Oracle\Java\javapath	Akses JDK agar bisa mendukung spark apache
C:\spark\scala\bin	Untuk menjalankan scala untuk menjalankan data frame, dan SQL
C:\spark\spark\bin	Agar spark dapat dioperasikan

Sumber : Hasil Penelitian (2021)

3.1.4. Membangun Cluster Komputer

Untuk menjalankan big data menggunakan spark apache adalah dengan mengakses pyspark. Untuk menjalankan pyspark. Berikut tahapan yang harus dilakukan:

1. Koneksi spark master

Masuk ke anaconda command prompt kemudian arahkan ke directory dimana spark apache/bin berada. Jalankan koneksi ke master slave sebagai computer slave dengan menggunakan perintah: `spark-class org.apache.spark.deploy.master.Master`.

2. Koneksi computer worker

Komputer worker digunakan sebagai client dimana data yang akan deprogram oleh python dilakukan di worker. Untuk masuk ke worker untuk koneksi ke worker: `spark-class org.apache.spark.deploy.worker.Worker spark://192.168.43.168:7077`. Ip 192.168.43.168:7077 adalah alamat dari master slave, dengan koneksi ke master slave big data yang disini menggunakan pyspark dapat dikoneksikan.

3. Menjalankan spark apache

Setelah master slave dan worker dijalankan maka pyspark dapat dijalankan. Memanggil pyspark adalah untuk masuk ke directory dimana data-data atau menjalankan python dilakukan

3.2. Implementasi Kmean Clustering menggunakan Spark Apache

3.2.1. Preprocessing

Tahapan ini adalah mempersiapkan data yang akan diproses memeriksa memilih, menghapus atau menghapus data yang akan digunakan dalam penelitian ini. Data diolah menggunakan Microsoft excel dan simpan dengan format xls. Data yang telah diolah simpan

didalam directory riset internal yang selanjutnya akan tampil dalam perangkat lunak Jupiter. Berikut listing program untuk persiapan input data dalam dilihat pada Gambar 4.7.

```

In [1]: #mengimport modul yang dibutuhkan
from pyspark.ml.clustering import KMeans
from pyspark.ml.feature import VectorAssembler

#membuat session
appName = "Klastering di spark"
spark = SparkSession \
    .builder \
    .appName(appName) \
    .config("spark.some.config.option", "some-value") \
    .getOrCreate()
    
```

Sumber : Hasil Penelitian (2021)

Gambar 2. Hasil instalasi Spark Apache

Untuk input data dari dataset yang sudah tersimpan di directory jupyter. Dimana ada beberapa dataset yang akan dikelompokan data menggunakan k-mean clustering. Berikut salah contoh salah satu input data dari excel kedalam jupyter menggunakan python, dapat dilihat pada gambar 3.

```

In [6]: #membuat data dari file ke DataFrame dengan infer skema
customers = spark.read.csv(
    "dataset/student_GPA.csv", inferSchema=True, header=True)
customers.show(106)
    
```

absences	G1	G2	G3
0	5	0	0
4	5	5	0
10	7	0	10
2	15	14	15
4	0	10	10
10	15	15	15
0	12	12	11
6	0	5	0
0	10	10	10
0	14	15	15
0	10	0	9
4	10	12	12
2	14	14	14
2	10	10	11
0	14	16	16
4	14	14	14

Sumber: Hasil Penelitian (2021)

Gambar 3. Proses Input Data

3.2.2. Data Training

Tahap selanjutnya adalah menyiapkan data training sebelum memproses K-mean clusteringnya. Dimana dalah tahap ini sebelum dimasukan ke model kmean data dirubah kedalam bentuk vector assembler. Berikut listing program tahap training menggunakan python dapat dilihat pada gambar 4.


```
In [12]: #membuat assembler untuk mengubah fitur menjadi satu kolom fitur
assembler = VectorAssembler(inputCols = [
    "absences", "G1", "G2", "G3"],
    outputCol="Features")
train = assembler.transform(customers).select('Features')
train.show(truncate = false, n=10)

+-----+
|Features|
+-----+
[[0.0,5.0,0.0,0.0]
 [4.0,5.0,5.0,0.0]
 [10.0,7.0,8.0,10.0]
 [2.0,15.0,14.0,15.0]
 [4.0,0.0,10.0,10.0]
 [10.0,15.0,15.0,15.0]
 [0.0,12.0,12.0,11.0]
 [6.0,6.0,5.0,6.0]
 [0.0,10.0,10.0,15.0]
 [0.0,14.0,15.0,15.0]
+-----+
only showing top 10 rows
```

Sumber : Hasil Penelitian (2021)

Gambar 4. List Program Untuk Data Training

3.2.3. Membuat Model K-Mean

Tahapan ini adalah proses membangun model k-mean clustering, menentukan jumlah cluster (kelompok) data yang akan digunakan untuk proses prediksi. Berikut listing program dapat dilihat pada gambar 5.

```
#mendefinisikan algoritma kclustering
kmeans = KMeans(
    featuresCol=assembler.getOutputCol(), predictionCol="cluster",
    k=5, seed=0)
#mentraining model dengan perintah ".fit()"
model = kmeans.fit(train)
```

Sumber : Hasil Penelitian (2021)

Gambar 5. Membangun Model K-mean

3.2.4. Menentukan Titik Cluster

Pada tahapan ini adalah menentukan titik pusat cluster (centroid), dimana dalam penelitian ini untuk menentukan nilai awal centroid dilakukan secara acak. Dalam penelitian ini ditentukan kluster (kelompok) sejumlah 5 cluster. Berikut implementasi penentuan titik awal cluster menggunakan Bahasa pemograman phyton, ditampilkan pada gambar 6.

```
In [4]: 1 #mendefinisikan algoritma kclustering
        2 kmeans = KMeans(
            3 featuresCol=assembler.getOutputCol(), predictionCol="prediction cluster",
            4 k=5, seed=0)
        5 #mentraining model dengan perintah ".fit()"
        6 model = kmeans.fit(train)
        7

In [20]: 1 centers = model.clusterCenters()
         2 print("Cluster Centers: ")
         3 for center in centers:
         4     print(center)

Cluster Centers:
[[0.45454545 7.31818182 4.79545455 0.77272727]
 [2.34013605 9.24489706 9.6123449 0.7416966 ]
 [ 3.04724489 14.48818889 14.58267717 14.82464567]
 [62.6 10.2 10. 0.4]
 [15.29446444 10.23611111 0.80533556 0.95833333]]
```

Sumber : Hasil Penelitian (2021)

Gambar 6. Listing Program Untuk Menentukan Centroid

3.2.5. Prediksi Cluster

Hasil terakhir dari tahapan pengelompokan data (clustering) dalam algoritma k-Mean adalah menentukan tiap record termasuk dalam cluster berapa. Dimana dalam penelitian ini sudah ditentukan jumlah cluster adalah 5. Berikut implementasi python untuk memprediksi cluster, dapat dilihat pada gambar 7.

```
In [30]: 1 prediction = model.transform(train)#melakukan prediksi kcluster
        2 prediction.groupBy("prediction cluster").count().orderBy("prediction cluster").show()
        3 prediction.select('no', 'prediction cluster').show(395)#menampilkan 5 data hasil prediksi
```

Sumber : Hasil Penelitian (2021)

Gambar 7. Listing Program Untuk Prediksi Cluster

Hasil yang didapatkan dari data record berjumlah 395 menunjukkan cluster 0 berjumlah 44, cluster 1 berjumlah 147, cluster 2 berjumlah 127, cluster 3 berjumlah 5 dan cluster 4 berjumlah 72. Lengkap dapat dilihat pada gambar 8.

prediction cluster	count
0	44
1	147
2	127
3	5
4	72

Sumber : Hasil Penelitian (2021)

Gambar 8. Hasil prediksi cluster

Hasil prediksi dari setiap record yang telah dilakukan proses clustering dengan k-mean dapat dilihat dari gambar 9.

```
In [8]: 1 df_pred = no.join(transformed, 'no')
        2 df_pred.show()
```

	No	absences	G1	G2	G3	prediction cluster
1	6	5	6	6	1	
2	4	5	5	6	0	
3	10	7	8	10	4	
4	2	15	14	15	2	
5	4	6	10	10	1	
6	10	15	15	15	2	
7	0	12	12	11	1	
8	0	6	5	0	1	
9	0	16	18	19	2	
10	0	14	15	15	2	
11	0	10	8	9	1	
12	4	10	12	12	1	
13	2	14	14	14	2	
14	2	10	10	11	1	
15	0	14	16	16	2	
16	4	14	14	14	2	

Sumber : Hasil Penelitian (2021)

Gambar 8. Hasil prediksi cluster

3.3. Hasil Analisis implementasi Big Data

Dalam penelitian ini menghasilkan pengelompokan data dari dataset student performance yang termasuk bagian dari analytical academic untuk analisis big data di perguruan tinggi. Apache Spark merupakan salah satu bagian dari proyek Big Data Apache Software Foundation yang menangani analisis data pada Big Data. Dimana untuk menggunakan k-mean clustering dalam memprediksi data membutuhkan library pyspark.ml.clustering import KMeans. Apache spark digunakan oleh python agar ada computer yang dapat digunakan sebagai server (Master Slave) dan sebagai client (worker) serta untuk menjalankan library Machine Learning atau disingkat dengan MLlib di python.

Hasil yang didapat dalam analisis big data ini, k-mean clustering berhasil melakukan prediksi menjadi 5 cluster sesuai dengan yang ditentukan pada centroid dan tidak terjadi kendala.

4. Kesimpulan

Dari hasil penelitian implemementasi analisis big data pada perguruan tinggi menggunakan machine learning disimpulkan bahwa Analisis big data pada perguruan tinggi meliputi learning analytical, academic analytical dan analytical staff. Dalam penelitian ini dilakukan pada bidang analytical academic dengan memprediksi student performance dengan dataset yang didapatkan dataset public. Dalam peneltian Analisis big data dioperasikan menggunakan apache spark selanjutnya untuk proses pengelompokan data menggunakan algoritma k-mean clustering bagian dari algoritma machine learning. Hasil kinerja mahasiswa didapat dari 4 variable yaitu nilai abseces, grade1, grade 2 dan grade 3 dalam dikelompokan kedalam 5 cluster. Dimulai dari cluster 1 termasuk nilai rata-rata yang paling besar, dan cluser ini adalah kelompok dari mahasiswa yang nilai terkecil.

Untuk pengembangan penelitian selanjutnya kriteria learning analytical, academic analytical, dan staff aniytical harus diolah dan dilakukan analis bersama-sama. Untuk hasil

kinerja big data lebih baik, sebaiknya menggunakan computer server atau cloud computing. Sebaiknya menggunakan dataset yang langung diambil dari system informasi akademik tiap perguruan tinggi untuk mendapatkan nilai maximal dalam penelitian

Ucapan Terima Kasih

Terima Kasih Untuk LPPM yang sudah Mendanai Penelitian internal sehingga penelitain big data ini bisa diselesaikan.

Daftar Pustaka

- ADMIN SEVIMA. (2019). Manfaat dan Penggunaan Big Data Analytic untuk Perguruan Tinggi. Retrieved from Sevima website: <https://sevima.com/manfaat-dan-penggunaan-big-data-analytic-untuk-perguruan-tinggi/>
- Asniar. (2015). *Penggunaan Big Data Analytic di Perguruan Tinggi*. (June), 1–5. <https://doi.org/10.13140/RG.2.1.4581.9046>
- Aurelia, R. (2017). *IMPLEMENTASI METODE K-MEANS CLUSTERING DALAM MENGELOMPOKKAN EMOSI SENANG, MARAH, DAN NETRAL BERDASARKAN VOKAL MANUSIA*. Universitas Multimedia Nusantara.
- Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. <https://doi.org/10.1111/bjet.12230>
- Jamshidi, M. M. (2017). A system of systems framework for autonomy with big data analytic and machine learning. *Procedia Computer Science*, 120, 6. <https://doi.org/10.1016/j.procs.2017.11.202>
- Kaur, P., Sharma, M., & Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. *Procedia Computer Science*, 132, 1049–1059. <https://doi.org/10.1016/j.procs.2018.05.020>
- Kumar Sinha, S., & Singh, H. (2019). Significance of Big Data and Analytics of Student Success in Higher Education. *Journal of Computer Science and Information Technology*, 8(11), 7–12. Retrieved from www.ijcsmc.com
- Murumba, J., & Micheni, E. (2017). Big Data Analytics in Higher Education: A Review. *The International Journal of Engineering and Science*, 06(06), 14–21. <https://doi.org/10.9790/1813-0606021421>
- Pratiwi, E. S., & Herlawati, H. (2019). Sistem Informasi Penjualan Katering Berbasis Web Pada CV. Saung Alit Telaga Murni Cikarang Barat Eka. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS*, 03(2), 177–188.
- Tulasi, B. (2013). Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 68(14), 21–23. <https://doi.org/10.5120/11648-7142>
- Tulasi, B. (2014). Learning Analytics and Big Data in Higher Education. *International Journal of Engineering Research & Technology (IJERT)*, 3(1), 3377–3383.

- Vinay, A., Shekhara, V. S., Rituparna, J., Aggrawal, T., Balasubramanya Murthy, K. N., & Natarajan, S. (2015). Cloud based big data analytics framework for face recognition in social networks using machine learning. *Procedia Computer Science*, 50, 623–630. <https://doi.org/10.1016/j.procs.2015.04.095>
- Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tourism Management*, 82(October 2019), 104208. <https://doi.org/10.1016/j.tourman.2020.104208>



Plagiarism Checker X Originality Report

Similarity Found: 25%

Date: Monday, September 06, 2021

Statistics: 726 words Plagiarized / 2960 Total words

Remarks: Medium Plagiarism Detected - Your Document needs Selective Improvement.

Journal of Information and Information Security (JIFORTY) Vol. 2, No. 1, Juni 2021, 77-88
e-ISSN : 2722-4058 Available Online at <http://ejurnal.ubharajaya.ac.id/index.php/jiforty>
77 **Implementasi Big Data Analytical Untuk Perguruan Tinggi** Menggunakan Machine Learning Rakhmat Purnomo^{1,*}, Wowon Priatna¹, Tri Dharma Putra¹ 1 Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Perjuangan 081, Marga Mulya, Bekasi Utara; 02188955882; e-mail: rakhmat.purnomo@dsn.ubharajaya.ac.id, wowon.priatna@dsn.ubharajaya.ac.id, tridharma.putra@dsn.ubharajaya.ac.id *
Korespondensi: e-mail: rakhmat.purnomo@dsn.ubharajaya.ac.id Diterima: 10 Juni 2021; Review: 26 Juni 2021; Disetujui: 29 Juni 2021; Diterbitkan: 3 Juli 2021 Abstract The dynamics of higher education are changing and emphasize the need to adapt quickly.

Higher education is under the supervision of accreditation agencies, governments and other stakeholders to seek new ways to improve and monitor student success and other institutional policies. Many agencies fail to make efficient use of the large amounts of available data. With the use of big data analytics in higher education, it can be obtained more insight into students, academics, and the process in higher education so that it supports predictive analysis and improves decision making.

The purpose of this research is to implement big data analytical to increase the decision making of the competent party. This research begins with the identification of process data based on analytical learning, academic and process in the campus environment. The data used in this study is a public dataset from UCI machine learning, from the 33 available variables, 4 variables are used to measure student performance.

Big data analysis in this study uses spark apace as a library to operate pyspark so that python can process big data analysis. The data already in the master slave is grouped

using k-mean clustering to get the best performing student group. The results of this study succeeded in grouping students into 5 clusters, cluster 1 including the best student performance and cluster 5 including the lowest student performance. Keywords: Machine Learning, Big Data Analytical, Learning Academic, k-mean Clustering Python, Hadoop, apache spark.

Abstrak Dinamika pendidikan tinggi sedang berubah dan menekankan kebutuhan untuk beradaptasi dengan cepat. Pendidikan tinggi berada di bawah pengawasan lembaga akreditasi, pemerintah, dan pemegang kepentingan lainnya untuk mencari cara baru untuk meningkatkan dan memantau keberhasilan mahasiswa dan kebijakan kelembagaan lainnya. Banyak lembaga gagal memanfaatkan sejumlah besar data yang tersedia secara efisien.

Dengan penggunaan big data analytic di perguruan tinggi maka dapat diperoleh wawasan yang lebih tentang mahasiswa, akademisi, dan proses di perguruan tinggi sehingga mendukung analisis prediksi dan peningkatan pengambilan keputusan. Tujuan penelitian ini adalah untuk implementasi big data analytical untuk meningkatkan pengambilan keputusan pihak kampus.

Penelitian ini dimulai dengan identifikasi data-data proses berdasarkan learning analytical, academic dan proses di lingkungan kampus. Data yang digunakan dalam penelitian ini adalah dataset public dari bersumber dari UCI machine learning, dari 33 variabel tersedia digunakan 4 variabel untuk mengukur kinerja mahasiswa.

Big data analisis dalam penelitian ini menggunakan spark apache sebagai library untuk mengoperasikan pyspark agar python dapat mengolah big data analisis. Data yang sudah ada di master slave dilakukan pengelompokan data menggunakan k-mean clustering untuk mendapatkan kelompok mahasiswa yang kinerja terbaik. Hasil dari penelitian ini Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 78 Jurnal JIFORTY 2 (1): 77-88(Juni 2021) berhasil mengelompokkan mahasiswa menjadi 5 cluster, cluster 1 termasuk dalam kinerja mahasiswa terbaik dan cluster 5 termasuk kinerja mahasiswa yang terendah.

Kata kunci: Machine Learning, Big Data Analytical, Learning Academic, k-mean Clustering Python, Hadoop, apache spark. 1. Pendahuluan Perkembangan Teknologi digital di berbagai ruang dan waktu terus menghasilkan data dalam jumlah besar. Big Data menggambarkan pertumbuhan yang signifikan dalam volume dan variasi data yang tidak dapat lagi dikelola menggunakan database tradisional.

Dengan bantuan analitik, jumlah data yang tampaknya berbeda dan heterogen ini dapat

diproses untuk pola, yang pada dapat menghasilkan wawasan penting untuk pengambilan keputusan (Daniel, 2015) Dampak Digitalisasi data telah membuka peluang penggunaan big data di perguruan tinggi. Nilai big data terletak pada hasil analisis dan prediksi atau tindakan yang diambil dari hasil analisis dan prediksi tersebut (ADMIN SEVIMA, 2019).

Dinamika pendidikan tinggi sedang berubah dan menekankan kebutuhan untuk beradaptasi dengan cepat. Pendidikan tinggi berada di bawah pengawasan lembaga akreditasi, pemerintah, dan pemegang kepentingan lainnya untuk mencari cara baru untuk meningkatkan dan memantau keberhasilan siswa dan kebijakan kelembagaan lainnya (Tulasi, 2013) Pendidikan masa depan lebih sering dihubungkan dengan teknologi baru, pendidikan tinggi beroperasi dengan lingkungan semakin kompleks (Asniar, 2015) seperti perangkat komputasi di mana-mana, desain ruang kelas yang fleksibel.

Karena institusi pendidikan tinggi mempunyai banyak data tentang mahasiswa, dan basis data catatan mahasiswa menjadi lebih kompleks untuk diakses, maka Perguruan Tinggi harus mendata semua data terkait akademik dari berbagai kegiatan seperti data mahasiswa, data registrasi, data asesmen dan lainnya. Untuk mempermudah pendidikan tinggi dalam mengakes yang dapat dihasilkan oleh mahasiswa adalah menerapkan analytical big data dalam e-learning analytical, dimana institusi dapat menentukan mahasiswa yang sukses atau mahasiswa yang drop out (Asniar, 2015).

Penelitian (Murumba & Micheni, 2017) menerapkan penerapan big data mengeksplorasi atribut big data yang relevan dengan institusi pendidikan. Penelitian (Tulasi, 2014) menerapkan big data analisis dalam analisa pembelajaran dengan menganalisa data statis dan dinamis dilingkungan belajar-mengajar untuk memungkinkan intervensi tepat waktu oleh pendidik sedangkan (Kumar Sinha & Singh, 2019) rekomendasi analisis big data dalam mencapai keberhasilan mahasiswa diberbagai perguruan tinggi. Penelitian (Daniel, 2015) menerapkan big data untuk rekomendasi project untuk proyek big data untuk tingkat kelembagaan institusi.

Beberapa Penelitian menggunakan kombinasi big data dan machine learning dalam menyelesaikan permasalahan data besar diantaranya: menggunakan machine learning dalam mengolah data besar untuk sisi keamanan (Kaur, Sharma, & Mittal, 2018) untuk memperkirakan Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning Copyright © 2021 Jurnal JIFORTY 2 (1): 77-88 (Juni 2021) 79 permintaan pariwisata kapal pesiar Tiongkok (Xie, Qian, & Wang, 2021) pengenalan wajah (Vinay et al.,

2015) kerangka sistem untuk otonomi (Jamshidi, 2017) Dari latar belakang dan penelitian tujuan penelitian ini adalah untuk implementasi big data analisis diperguruan tinggi dengan berfokus kepada learning academic dalam menentukan prediksi kinerja mahasiswa menggunakan k-mean clustering. 2. Metode Penelitian Dalam Penelitian ini menggunakan dataset yang bersifat public dapat Students Academic Performance yang bersumber dari <https://archive.ics.uci.edu> jumlah 32 atribut diambil 4 yang akan digunakan dalam variable yang akan diolah oleh k-mean clustering. Tahapan dalam penelitian ini dapat dilihat pada gambar 1.

Sumber : Hasil Penelitian (2021) Gambar 1. Tahapan Penelitian 2.1. Input Data Data yang telah diperoleh berupa format excel akan dirubah dan disesuaikan dengan format csv yang akan digunakan pada pemograman python, selanjutnya akan diinput kedalam directory folder yang nantinya akan ditampilkan pada directory Jupiter.

Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 80 Jurnal JIFORTY 2 (1): 77-88(Juni 2021) 2.2. Data Preprosesing Dalam tahapan persiapan pemrosesan data adalah mengecek data, memastikan tidak record, attribute yang kosong atau pun format sudah sesuai yang akan diolah python.

Data yang akan diolah adalah data student performace sebagai bagian dari academic analytical dalam penerapan big data analytical dalam perguruan tinggi. 2.3. Spark Apache Dalam tahapan ini adalah instalasi spark apache dan perangkat lunak lainnya. Spark apache ini berfungsi untuk mengoperasikan big data analisis. Dimana beberapa software pendukung yang harus diinstall adalah: Anaconda ,Spark apache, Hadoop,Java development kit (JDK) dan skala.

Dalam tahapan ini setelah semua perangkat lunak telah terinstall dan spark apache telah berhasil digunakan. Selanjutnya akan membuat server dalam spark apache yang dinamakan master slave. Master slave yang telah berhasil dikoneksikan selanjutnya IP address akan digunakan untuk membuat koneksi worker (client). 2.4.

Data Selection (pemilihan Atribut) Data set dengan format csv dapat diproses dengan Bahasa pemrograman Python. Maka, rentang data yang akan diolah harus didefinisikan terlebih dahulu. melakukan seleksi rentang data adalah Impor fungsi arrange dari library numpy serta membuat sebuah variabel yang menampung fungsi arrange Definiskan secara numerik jumlah baris data yang ingin diseleksi Proses ini akan menghasilkan variabel yang menampung data yang telah diseleksi. 2.5.

K-mean Clustering Untuk menjalankan big data analisis dengan Algoritma k-mean Clustering dipemograman python dibutuhkan beberapa tahapan: 1. Mengimport

pyspark (sebuah library big data menggunakan spark apache 2. Mengimport library machine learning (MLlib) yang dibutuhkan oleh python untuk memproses machine learning 3. Mengimport library k-mean clustering 4.

Menentukan **centroid (titik pusat cluster)** menggunakan mean masing-masing kelompok data 5. Mealokasikan data kepada centroid terdekat 6. Proses berulang Melakukan iterasi untuk memastikan tidak ada data yang pindah cluster. 2.6. Evaluasi Kinerja Pada tahapan ini akan menguji kinerja dari hasil prediksi k-mean clustering apakah sudah efektif sehingga dapat digunakan sebagai rekomendasi model untuk digunakan.

Implementasi **Big Data Analytical Untuk Perguruan Tinggi** Menggunakan Machine Learning Copyright © 2021 Jurnal JIFORTY 2 (1): 77-88 (Juni 2021) 81 **3. Hasil dan Pembahasan 3.1.** Perancangan Big Data 3.1.1. Instalasi Spark apache Dalam penelitian ini untuk menjalankan Big Data digunakan spark apache untuk memproses data berskala besar.

Spark apache menyediakan library untuk mengoperasikan python, distribusi skala, dan menampung Hadoop. Spark apache yang digunakan adalah versi 2.47 digunakan untuk membuat master Komputer sebagai server dan mengoperasikan akses client menggunakan worker. Berikut tampilan spark apache yang sudah terinstall di computer **dapat dilihat pada gambar** 2. Sumber : Hasil Penelitian (2021) Gambar 2.

Hasil instalasi Spark Apache 3.1.2. Instalasi apache Hadoop Apache Hadoop merupakan salah satu distribusi big data yang dikembangkan oleh Apache Software Foundation. Apache Hadoop ini dibangun dengan lisensi free dan open source. Dalam penelitian ini digunakan Apache Hadoop sebagai framework untuk Big Data dengan versi 2.7.0. Instalasi dan konfigurasi dilakukan pertama kali **pada komputer yang bertindak sebagai** hadoop-master (single node), untuk menjadikan multimode cluster akan dilakukan sinkronisasi dengan komputer slave dan dengan melakukan perubahan konfigurasi yang minimal pada node.

Secara garis besar tahapan instalasi dan konfigurasi Big Data Hadoop dengan Install Java Development Kit (JDK), Buat user dan group untuk Hadoop, Konfigurasi firewall dinonaktifkan dan Instalasi Hadoop dan konfigurasi Hadoop Environment pada node master dan slave. Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 82 Jurnal JIFORTY 2 (1): 77-88(Juni 2021) 3.1.3.

Konfigurasi Environment Pada tahapan ini adalah setting environment variable pada my computer agar big data dapat digunakan. Setting environment **dapat dilihat pada tabel** 1. Tabel 1. Setting Environment Variable Fungsi PYSPARK_DRIVER_PYTHON =

E:\phyton\scripts\jupyter.exe Untuk menjalankan python pada spark apache. Dimana editor menggunakan jupyter.exe PYSPARK_DRIVER_PYTHON_OPTS = notebook Tempat menulis code PYSPARK_PYTHON = E:\phyton\python.exe Agar pyspark dapat dijalankan C:\Program Files (x86)\Common Files\Oracle\Java\javapath Akses JDK agar bisa mendukung spark apache C:\spark\scala\bin Untuk menjalankan scala untuk menjalankan data frame, dan SQL C:\spark\spark\bin Agar spark dapat dioperasikan Sumber : Hasil Penelitian (2021) 3.1.4.

Membangun Cluster Komputer Untuk menjalankan big data menggunakan spark apache adalah dengan mengakses pyspark. Untuk menjalankan pyspark. Berikut tahapan yang harus dilakukan: 1. Koneksi spark master Masuk ke anaconda command prompt kemudian arahkan ke directory dimana spark apache/bin berada. Jalankan koneksi ke master slave sebagai computer slave dengan menggunakan perintah: spark-class org.apache.spark.deploy.master.Master. 2.

Koneksi computer worker Komputer worker digunakan sebagai client dimana data yang akan deprogram oleh python dilakukan di worker. Untuk masuk ke worker untuk koneksi ke worker: spark-class org.apache.spark.deploy.worker.Worker spark://192.168.43.168:7077. Ip 192.168.43.168:7077 adalah alamat dari master slave, dengan koneksi ke master slave big data yang disini menggunakan pyspark dapat dikoneksikan. 3.

Menjalankan spark apache Setelah master slave dan worker diajalankan maka pyspark dapat dijalankan. Memanggil pyspark adalah untuk masuk ke directory dimana data-data atau menjalankan python dilakukan 3.2. Implementasi Kmean Clustering menggunakan Spark Apache 3.2.1. Preprosesing Tahapan ini adalah mempersiapkan data yang akan diproses memeriksa memilih, menghapus atau menghapus data yang akan digunakan dalam penelitian ini.

Data diolah menggunakan Microsoft excel dan simpan dengan format xls. Data yang telah diolah simpan Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning Copyright © 2021 Jurnal JIFORTY 2 (1): 77-88 (Juni 2021) 83 didalam directory riset internal yang selanjutnya akan tampil dalam perangkat lunak Jupiter.

Berikut listing program untuk persiapan input data dalam dilihat pada Gambar 4.7. Sumber : Hasil Penelitian (2021) Gambar 2. Hasil instalasi Spark Apache Untuk input data dari dataset yang sudah tersimpan di directory jupyter. Dimana ada beberapa dataset yang akan dikelompokan data menggunakan k-mean clustering.

Berikut salah satu contoh input data dari excel kedalam jupyter menggunakan python, **dapat dilihat pada gambar 3**. Sumber: Hasil Penelitian (2021) Gambar 3. Proses Input Data 3.2.2. Data Training Tahap selanjutnya adalah menyiapkan data training sebelum memproses K-mean clusteringnya. Dimana dalam tahap ini sebelum dimasukkan ke model kmean data dirubah kedalam bentuk vector assembler.

Berikut listing program tahap training menggunakan python **dapat dilihat pada gambar 4**. Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 84 Jurnal JIFORTY 2 (1): 77-88(Juni 2021) Sumber : Hasil Penelitian (2021) Gambar 4. List Program Untuk Data Training 3.2.3.

Membuat Model K-Mean Tahapan ini adalah proses membangun model k-mean clustering, menentukan jumlah cluster (kelompok) data **yang akan digunakan untuk** proses prediksi. Berikut listing program **dapat dilihat pada gambar 5**. Sumber : Hasil Penelitian (2021) Gambar 5. Membangun Model K-mean 3.2.4. Menentukan Titik Cluster Pada tahapan ini adalah **menentukan titik pusat cluster** (centroid), dimana dalam penelitian ini untuk menentukan nilai awal centroid dilakukan secara acak. Dalam penelitian ini ditentukan kluster (kelompok) sejumlah 5 cluster.

Berikut implementasi penentuan titik awal cluster menggunakan Bahasa pemograman python, ditampilkan pada gambar 6. Sumber : Hasil Penelitian (2021) Gambar 6. Listing Program Untuk Menentukan Centroid **Implementasi Big Data Analytical Untuk Perguruan Tinggi** Menggunakan Machine Learning Copyright © 2021 Jurnal JIFORTY 2 (1): 77-88 (Juni 2021) 85 3.2.5.

Prediksi Cluster Hasil terakhir dari tahapan pengelompokan data (clustering) dalam algoritma k-Mean adalah menentukan tiap record termasuk dalam cluster berapa. Dimana dalam penelitian ini sudah ditentukan jumlah cluster adalah 5. Berikut implementasi python untuk memprediksi cluster, **dapat dilihat pada gambar 7**. Sumber : Hasil Penelitian (2021) Gambar 7.

Listing Program Untuk Prediksi Cluster Hasil yang didapatkan dari data record berjumlah 395 menunjukkan cluster 0 berjumlah 44, cluster 1 berjumlah 147, cluster 2 berjumlah 127, cluster 3 berjumlah 5 dan cluster 4 berjumlah 72. Lengkap **dapat dilihat pada gambar 8**. Sumber : Hasil Penelitian (2021) Gambar 8. Hasil prediksi cluster Hasil prediksi dari setiap record yang telah dilakukan proses clustering dengan k-mean dapat dilihat dari gambar 9.

Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 86 Jurnal JIFORTY 2 (1): 77-88(Juni 2021) Sumber : Hasil Penelitian (2021) Gambar 8. Hasil prediksi cluster 3.3. Hasil Analisis

implementasi Big Data Dalam penelitian ini menghasilkan pengelompokan data dari dataset student performance yang termasuk bagian dari analytical academic untuk analisis big data di perguruan tinggi.

Apache Spark merupakan salah satu bagian dari proyek Big Data Apache Software Foundation yang menangani analisis data pada Big Data. Dimana untuk menggunakan k-mean clustering dalam memprediksi data membutuhkan library pyspark.ml.clustering import KMeans. Apache spark digunakan oleh python agar ada computer yang dapat digunakan sebagai server (Master Slave) dan sebagai client (worker) serta untuk menjalankan library Machine Learning atau disingkat dengan MLlib di python.

Hasil yang didapat dalam analisis big data ini, k-mean clustering berhasil melakukan prediksi menjadi 5 cluster sesuai dengan yang ditentukan pada centroid dan tidak terjadi kendala. 4. Kesimpulan Dari hasil penelitian implementasi analisis big data pada perguruan tinggi menggunakan machine learning disimpulkan bahwa Analisis big data pada perguruan tinggi meliputi learning analytical, academic analytical dan analytical staff.

Dalam penelitian ini dilakukan pada bidang analytical academic dengan memprediksi student performance dengan dataset yang didapatkan dataset public. Dalam penelitian Analisis big data dioperasikan menggunakan apache spark selanjutnya untuk proses pengelompokan data menggunakan algoritma k-mean clustering bagian dari algoritma machine learning. Hasil kinerja mahasiswa didapat dari 4 variable yaitu nilai absences, grade1, grade 2 dan grade 3 dalam dikelompokkan kedalam 5 cluster.

Dimulai dari cluster 1 termasuk nilai rata-rata yang paling besar, dan cluster ini adalah kelompok dari mahasiswa yang nilai terkecil. Untuk pengembangan penelitian selanjutnya kriteria learning analytical, academic analytical, dan staff analytical harus diolah dan dilakukan analisis bersama-sama. Untuk hasil Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning Copyright © 2021 Jurnal JIFORTY 2 (1): 77-88 (Juni 2021) 87 kinerja big data lebih baik, sebaiknya menggunakan computer server atau cloud computing.

Sebaiknya menggunakan dataset yang langsung diambil dari system informasi akademik tiap perguruan tinggi untuk mendapatkan nilai maximal dalam penelitian Ucapan Terima Kasih Terima Kasih Untuk LPPM yang sudah Mendanai Penelitian internal sehingga penelitian big data ini bisa diselesaikan. Daftar Pustaka ADMIN SEVIMA. (2019). Manfaat dan Penggunaan Big Data Analytic untuk Perguruan Tinggi. Retrieved from Sevima website: <https://sevima.com/manfaat-dan-penggunaan-big-data-analytic-untuk-perguruan-tinggi/> Asniar. (2015). Penggunaan Big Data Analytic di

Perguruan Tinggi. (June), 1 – 5. <https://doi.org/10.13140/RG.2.1.4581.9046> Aurelia, R. (2017). IMPLEMENTASI METODE K-MEANS CLUSTERING DALAM MENGELOMPOKKAN EMOSI SENANG, MARAH, DAN NETRAL BERDASARKAN VOKAL MANUSIA.

Universitas Multimedia Nusantara. Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904 – 920. <https://doi.org/10.1111/bjet.12230> Jamshidi, M. M. (2017). A system of systems framework for autonomy with big data analytic and machine learning. *Procedia Computer Science*, 120, 6. <https://doi.org/10.1016/j.procs.2017.11.202> Kaur, P., Sharma, M., & Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. *Procedia Computer Science*, 132, 1049 – 1059. <https://doi.org/10.1016/j.procs.2018.05.020> Kumar Sinha, S., & Singh, H. (2019). Significance of Big Data and Analytics of Student Success in Higher Education.

Journal of Computer Science and Information Technology, 8(11), 7 – 12. Retrieved from www.ijcsmc.com Murumba, J., & Micheni, E. (2017). Big Data Analytics in Higher Education: A Review. *The International Journal of Engineering and Science*, 06(06), 14 – 21. <https://doi.org/10.9790/1813-0606021421> Pratiwi, E. S., & Herlawati, H. (2019). Sistem Informasi Penjualan Katering Berbasis Web Pada CV. Saung Alit Telaga Murni Cikarang Barat Eka.

INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS, 03(2), 177 – 188. Tulasi, B. (2013). Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 68(14), 21 – 23. <https://doi.org/10.5120/11648-7142> Tulasi, B. (2014). Learning Analytics and Big Data in Higher Education. *International Journal of Engineering Research & Technology (IJERT)*, 3(1), 3377 – 3383.

Rakhmat Purnomo, Wowon Priatna, Tri Dharma Putra 88 *Jurnal JIFORTY 2* (1): 77-88(Juni 2021) Vinay, A., Shekhara, V. S., Rituparna, J., Aggrawal, T., Balasubramanya Murthy, K. N., & Natarajan, S. (2015). Cloud based big data analytics framework for face recognition in social networks using machine learning. *Procedia Computer Science*, 50, 623 – 630. <https://doi.org/10.1016/j.procs.2015.04.095> Xie, G., Qian, Y., & Wang, S. (2021).

Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tourism Management*, 82(October 2019), 104208. <https://doi.org/10.1016/j.tourman.2020.104208>

INTERNET SOURCES:

<1% -

<http://103.28.220.26/index.php?page=1&ipp=10&ref=browse&mod=viewjournal&journal=149>

<1% - <https://core.ac.uk/download/pdf/322456381.pdf>

<1% -

<https://www.planetcompliance.com/the-future-is-here-and-businesses-need-to-adapt-quickly/>

6% -

https://www.researchgate.net/publication/271070362_Significance_of_Big_Data_and_Analytics_in_Higher_Education

2% -

https://www.researchgate.net/publication/317799370_Big_Data_Analytics_in_Higher_Education_A_Review

<1% -

<https://www.emerald.com/insight/content/doi/10.1108/AAOUJ-01-2017-0009/full/html>

6% -

https://www.researchgate.net/publication/277906196_Penggunaan_Big_Data_Analytic_diperguruan_Tinggi

1% -

<https://sevima.com/manfaat-dan-penggunaan-big-data-analytic-untuk-perguruan-tinggi/>

1% -

https://www.researchgate.net/publication/273635616_Learning_Analytics_and_Big_Data_in_Higher_Education

1% -

https://www.researchgate.net/publication/343748367_Forecasting_Chinese_cruise_tourism_demand_with_big_data_An_optimized_machine_learning_approach

<1% -

https://www.researchgate.net/publication/276291907_Cloud_Based_Big_Data_Analytics_Framework_for_Face_Recognition_in_Social_Networks_Using_Machine_Learning

1% -

https://www.researchgate.net/publication/325664068_Big_Data_and_Machine_Learning_Based_Secure_Healthcare_Framework

<1% -

https://www.researchgate.net/publication/351188622_Penanganan_Missing_Values_Untuk_Meningkatkan_Kinerja_Model_Machine_Learning_Pada_Data_Telemarketing

<1% -

<https://boyfandd1993.blogspot.com/2014/09/implementasi-algoritma-k-means-dalam.html>

<1% -

<http://fe.unj.ac.id/wp-content/uploads/2021/04/Buku-RIP-UNJ-2021-2025.pdf>

<1% -

<https://widuri.raharja.info/index.php?title=SI1411481295>

<1% -

http://cicilia_el.staff.gunadarma.ac.id/Downloads/files/44352/Konsep+dasar+pengembangan+sistem.pdf

<1% - <http://rizalespe.lecture.ub.ac.id/files/2015/10/DM-03-3-K-Means.pdf>

<1% - <http://seminar.uad.ac.id/index.php/STEEEM/article/download/2878/805>

<1% -

https://www.researchgate.net/publication/315989426_Perancangan_Papan_Informasi_Digital_Berbasis_Web_pada_Raspberry_pi

<1% -

<https://adoc.pub/daftar-isi-2-sistem-benam-single-board-computer-sbc-spesifik.html>

<1% - <http://ejournal.upnjatim.ac.id/index.php/scan/article/download/635/509>

<1% - https://dbgkiditanahpaua.wordpress.com/2015/02/26/admin_db/

<1% - <https://journal.unimal.ac.id/miej/article/download/375/269>

<1% -

<https://predatech.org/menentukan-keputusan-dengan-algoritma-klasifikasi-decision-tree-menggunakan-python/>

<1% -

<https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/download/598/pdf>

<1% - <https://jurnal.stmikroyal.ac.id/index.php/jurteksi/article/download/1060/630>

<1% - <http://tip.ppj.unp.ac.id/index.php/tip/article/download/55/40/>

<1% -

https://www.researchgate.net/publication/326894133_ANALISIS_CLUSTER_JARAK_ANTAR_KECAMATAN_DENGAN_MENGGUNAKAN_METODE_EUCLIDEAN_DI_UNTUK_PENETAPAN_ZONA_PENGOPERASIAN_MOBIL_MPLIK_DI_KOTA_MEDAN

<1% - <https://123dok.com/document/z1ekrk8y-atlas-rotan-indonesia-jilid.html>

<1% - <https://idbigdata.com/official/glossary/>

<1% -

https://www.researchgate.net/publication/332874818_Komparasi_Metode_Machine_Learning_Pada_Prediksi_Kelulusan_Mahasiswa

1% - <https://kc.umn.ac.id/view/divisions/Prodi=5FTI/2017.html>

<1% -

<https://www.studocu.com/en-gb/document/durham-university/educational-research-methods/bd-and-ed-core-reading-notes/16822853>

<1% - <https://www.sciencedirect.com/science/article/pii/S1877050917324122>

<1% -

https://www.researchgate.net/publication/335714660_An_Overview_of_Speech_Recognition_Using_HMM

<1% - <https://pathofscience.org/index.php/ps/article/view/620>

<1% - <https://link.springer.com/article/10.1007/s40593-021-00254-2>

<1% -

<https://www.springerprofessional.de/en/reviewing-and-exploring-innovative-ubiquitous-learning-tools-in-/16440210>

<1% - <https://ojs.unm.ac.id/ELT/article/view/10515>

<1% - <https://www.intechopen.com/profiles/46941>

<1% - <https://www.sciencedirect.com/science/article/pii/S0160738321000207>



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER

Kampus I: Jl. Harsono RM No.67, Ragunan, Pasar Minggu, Jakarta Selatan 12550
Telepon: (021) 27808121 - 27808882
Kampus II: Jl. Raya Perjuangan, Bekasi Utara, Jawa Barat
Telepon: (021) 88955882 Fax.: (021) 88955871
Web: www.ubharajaya.ac.id/fasilkom/. Email: fasikom@ubharajaya.ac.id

SURAT TUGAS

Nomor : ST/052/VI/2021/B-FASILKOM-UBJ

1. Dasar: Kalender Akademik Ubhara Jaya Tahun Akademik 2020/2021.
2. Dalam rangka mewujudkan Tri Dharma Perguruan Tinggi untuk Dosen di Universitas Bhayangkara Jakarta Raya maka dihimbau untuk melakukan penelitian.
3. Sehubungan dengan hal tersebut diatas, maka Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya menugaskan:

No.	NAMA	NIDN	JABATAN	KETERANGAN
1.	Rakhmat Purnomo, S.Pd., S.Kom., M.Kom	0322108201	Dosen Tetap Prodi Informatika	Sebagai Penulis Pertama
2.	Wowon Priatna, S.T., M.T.I	0429118707	Dosen Tetap Prodi Informatika	Sebagai Penulis Kedua
3.	Tri Dharma Putra, S.T., M.Sc	0302117101	Dosen Tetap Prodi Informatika	Sebagai Penulis Ketiga

Membuat Artikel Ilmiah dengan judul '**Implementasi Big Data Analytical Untuk Perguruan Tinggi Menggunakan Machine Learning**' pada Media Journal of Information and Information Security (JIFORTY), Vol. 2, No. 1, Juni 2021, Hal. 77-88, E-ISSN : 2722-4058.

4. Demikian penugasan ini agar dapat dilaksanakan dengan penuh rasa tanggung jawab.

Bekasi, 05 Juni 2021

DEKAN FAKULTAS ILMU KOMPUTER



HERLAWATI, S.Si., M.M., M.Kom.

NIP: 2001452

Paraf:

1. Ka. Prodi Informatika ..