

LAPORAN HASIL PENELITIAN



IMPLEMENTASI BIG DATA ANALYTICAL UNTUK PERGURUAN TINGGI MENGGUNAKAN MACHINE LEARNING

TIM PENELITIAN

TRI DHARMA PUTRA, S.T., M.Sc.

NIDN : 0302117101

RAKHMAT PURNOMO, S.Pd., S.Kom., M.Kom

NIDN : 0322108201

WOWON PRIATNA, ST, M.TI.

NIDN : 0429118007

**PROGRAM STUDI INFORMATIKA FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
MARET 2021**

LEMBAR PENGESAHAN LAPORAN HASIL PENELITIAN

1. Judul Penelitian : Implementasi Big Data Analytical untuk Perguruan Tinggi Menggunakan Machine Learning
2. Bidang Ilmu : Informatika.
3. Ketua Peneliti
 - a. Nama Lengkap dan Gelar : **Tri Dharma Putra, ST., M.Sc**
 - b. Jenis Kelamin : Laki - laki
 - c. NIDN : 0302117101
 - d. Jabatan Fungsional : Asisten Ahli
 - e. Fakultas/Program Studi : Ilmu Komputer/ Informatika
4. Jumlah Anggota Penelitian : 2
 - a. Anggota Peneliti 1 : Rakhmat Purnomo, S.Pd., S.Kom., M.Kom
 - b. Anggota Peneliti 2 : Wowon Priatna, ST, M.TI.
5. Lokasi Penelitian : -
6. Kerjasama dengan Institusi lain : -
7. Jangka waktu pelaksanaan : 3 bulan
8. Biaya : Rp. 4.000.000,-
Sumber dari Universitas : Rp. 4.000.000,-
9. Luaran yang dihasilkan : Jurnal Nasional ber ISSN (Nasional)

Jakarta, 31 Maret 2021

Mengetahui,
Dekan Fakultas Ilmu Komputer

Ketua Peneliti

Herlawati, S.Si., M.M., M.Kom.
NIP. 2001452

Tri Dharma Putra, ST., M.Sc
NIDN. 0302117101

Mengetahui,
Kepala Lembaga Penelitian dan Pengabdian Kepada Masyarakat dan Publikasi

Ir. Djuni Thamrin, M.Sc., Ph.D
NIP. 1908430

LEMBAR PENGESAHAN LAPORAN HASIL PENELITIAN

1. Judul Penelitian : Implementasi Big Data Analytical untuk perguruan tinggi menggunakan machine learning
2. Bidang Ilmu : Informatika.
3. Ketua Tim Pengusul
 - a. Nama Lengkap dan Gelar : **Tri Dharma Putra, ST., M.Sc**
 - b. Jenis Kelamin : Laki - laki
 - c. NIDN : 0302117101
 - d. Jabatan Fungsional : Asisten Ahli
 - e. Fakultas/Program Studi : Ilmu Komputer/ Informatika
4. Jumlah Anggota Penelitian : 2
 - a. Anggota Peneliti 1 : Rakhmat Purnomo, S.Pd., S.Kom., M.Kom
 - b. Anggota Peneliti 2 : Wowon Priatna, ST, M.TI.
5. Lokasi Penelitian : -
6. Kerjasama dengan Institusi lain : -
7. Jangka waktu pelaksanaan : 3 bulan
8. Biaya : Rp. 4.000.000,-
Sumber dari Universitas : Rp. 4.000.000,-
9. Luaran yang dihasilkan : Jurnal Nasional ber ISSN (Nasional)

Jakarta, 31 Maret 2021

Mengetahui,
Ka. Biro Keuangan

Ketua Peneliti

Tri Yulaeli, S.Pd., M.Ak., Akt.
NIP. 2101489

Tri Dharma Putra, ST., M.Sc
NIDN. 0302117101

Mengetahui,
Kepala Lembaga Penelitian dan Pengabdian Kepada Masyarakat dan Publikasi

Ir. Djuni Thamrin, M.Sc., Ph.D.
NIP. 1908430

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 siswa 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 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 mendapatkan 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

KATA PENGANTAR

Segala puji hanya bagi Allah, Tuhan Yang Maha Esa. Sholawat dan keselamatan untuk Nabi Muhammad, Rosul Akhir zaman. Atas ijin NYA lah penelitian ini dapat diselesaikan. Terima kasih kepada tim peneliti yang telah berkontribusi aktif dalam menyelesaikan penelitian ini.

Ucapan terima kasih kami sampaikan juga kepada,

1. Ibu Herlawati, S.Si, M.M., M.Kom, selaku dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jaya yang telah membantu mendorong dalam pelaksanaan Pengabdian Kepada Masyarakat.
2. Bapak Ir. Djuni Thamrin, MSc., PhD, selaku Ketua Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Bhayangkara Jaya yang telah memberi izin untuk melaksanakan Pengabdian Kepada Masyarakat di Masjid Asy-Syuhada.
3. Semua pihak yang tidak dapat kami sebutkan satu per satu dengan tidak mengurangi rasa hormat kami ucapkan terima kasih.

Kami menyadari banyak kekurangan dalam penelitian ini. Masukan dari rekan sejawat dan reviewer sangat kami nantikan.

Terima kasih.

Bekasi, 25 Februari 2021
Peneliti

DAFTAR ISI

LEMBAR PENGESAHAN LAPORAN HASIL PENELITIAN	i
LEMBAR PENGESAHAN LAPORAN HASIL PENELITIAN	ii
ABSTRAK.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2. Identifikasi Masalah	2
1.3. Rumusan Masalah	2
1.4. Tujuan Penelitian.....	2
1.5. Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Big Data.....	4
2.2. Machine Learning	4
2.3. K-mean clustering	4
2.3. Apache Spark	5
2.4. Penelitian Terdahulu	6
BAB III TUJUAN DAN MANFAAT PENELITIAN	9
3.1. Tujuan Penelitian.....	9
3.2. Manfaat Peneliti	9
BAB IV METODE PENELITIAN	10
3.1 Kerangka Pikir	10

3.2 Tahapan Penelitian.....	11
3.2.1. Input Data.....	11
3.2.2. Data Preprosesing	11
3.2.3. Spark apache	11
3.2.4 Data selection (pemilihan Atribut)	11
3.2.5. K-mean Clustering	12
3.2.6. Evaluasi Kinerja.....	12
3.3 Metode Pengumpulan data.....	12
3.3 Metode Analisis	14
BAB V IMPLEMENTASI.....	15
4.1 Spesifikasi Kebutuhan System.....	15
4.1.1 Spesifikasi Perangkat Keras.....	15
4.1.2 Spesifikasi Perangkat Lunak.....	15
4.2 Perancangan Big Data	16
4.2.1 Install Spark Apache	16
4.2.2 Instalasi apache Hadoop	16
4.2.3 Konfigurasi Environment.....	17
4.2.4 Membangun Cluster Komputer.....	17
4.3 Implementasi Kmean Clustering menggunakan Spark Apache	21
4.3.1 Preprocessing	21
4.3.2 Data Training	22
4.3.3 Membuat Model K-mean Clustering	22
4.3.4 Menentukan Titik Cluster	23
4.3.5 Prediksi Cluster	23
4.4 Hasil Analisis implementasi Big Data	24
BAB VI KESIMPULAN	25
5.1 Kesimpulan.....	25

5.2	Saran.....	25
REFERENSI.....		26
LAMPIRAN-LAMPIRAN		28
Lampiran 1. Dukungan sarana dan prasarana penelitian		28
Lampiran 2. Susunan organisasi tim peneliti dan pembagian tugas		29
Lampiran 3. Biodata Ketua dan Anggota		30
Lampiran 4. Surat Pernyataan Ketua Peneliti		35
Lampiran 5. Rincian Anggaran.....		36
Lampiran 6. Bukti Publikasi (in Review)		37
Lampiran 7. Log Book Penelitian.....		38

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu	6
Tabel 2. Informasi Data Student Performance	13
Tabel 3. Spesifikasi Hardware	15
Tabel 4. Spesifikasi Software	15
Tabel 5. Setting Environment	17

DAFTAR GAMBAR

Gambar 1. Apache Spark	5
Gambar 2. Kerangka Penelitian	10
Gambar 3. Tampilan Spark Apache	16
Gambar 4. Tampilan koneksi Master Slave	18
Gambar 5 Tampilan spark master	18
Gambar 6. Tampilan koneksi worker ke master slave	19
Gambar 7. Tampilan spark master deteksi worker	19
Gambar 8 Tampilan Running Aktivitas	20
Gambar 9. Menjalankan Pyspark	20
Gambar 10. Directory Jupiter	21
Gambar 11. list program untuk import library	21
Gambar 12. list program untuk import library	22
Gambar 13 list program untuk Data Training	22
Gambar 14 Membangun Model K-mean	22
Gambar 15. Listing Program Untuk Menentukan Centroid.....	23
Gambar 16. Listing Program Untuk Prediksi	23
Gambar 17. Hasil prediksi cluster.....	24
Gambar 18. Hasil Prediksi Tiap Record	24

DAFTAR LAMPIRAN

Lampiran 1. Dukungan sarana dan prasarana penelitian	28
Lampiran 2. Susunan organisasi tim peneliti dan pembagian tugas	29
Lampiran 3. Biodata Ketua dan Anggota	30
Lampiran 4. Surat Pernyataan Ketua Peneliti	35
Lampiran 5. Rincian anggaran	36
Lampiran 6. Bukti Publikasi (Submit Artikel).....	37

BAB I

PENDAHULUAN

1.1 Latar Belakang

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. 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 (B.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 mengakses 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 kelembangaan 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 et al., 2018), untuk memperkirakan permintaan pariwisata kapal pesiar Tiongkok (Zhang et al., 2020), pengenalan wajah (Vinay et al., 2015), kerangka sistem untuk otonomi (Jamshidi, 2017).

Dari latar belakang dan penelitian diatas maka penulis bermaksud untuk menerapkan big data analytical berdasarkan data yang didapat diperguruan tinggi kemudian di modelkan menggunakan machine learning sebagai pembelajaran machine.

1.2. Identifikasi Masalah

Adapun identifikasi masalah dalam penelitian ini adalah:

1. Banyak perguruan tinggi menghadapi masalah kompleks dalam lingkungan digitalisasi saat ini
2. Intitusi pendidikan tinggi masih banyak yang belum menyadari dalam mengolah data besar
3. Institusi pendidikan masih kesulitan dalam menentukan keberhasilan dan kegagalan mahasiswa dari data aktivitas mahasiswa.

1.3. Rumusan Masalah

Adapun Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi big data berdasarkan academic analitical?
2. Bagaimana memprediksi kinerja Mahasiswa menggunakan K-mean clustering

1.4. Tujuan Penelitian

Adapun Tujuan dalam penelitian ini adalah:

1. Memberikan usulan penggunaan big data analisis untuk perguruan tinggi berdasarkan academic analitical
2. Untuk kinerja siswa berdasarkan pengolompokan data
3. Untuk memprediksi keberhasilan mahasiswa dalam pembelajaran berdasarkan implementasi big data

1.5. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Analisis big data focus pada registrasi perkuliahan, pembayaran, partisipasi di kelas, belajar online, dan penilaian.
2. Tool yang digunakan Hadoop untuk mengolah data
3. Untuk algoritma machine learning diolah menggunakan 1 algoritma

BAB II

TINJAUAN PUSTAKA

2.1 Big Data

Big data merupakan kumpulan data yang memiliki ukuran besar dan melebihi kapasitas dari perangkat lunak basis data untuk mengelola dan menganalisanya. Big data membutuhkan teknik dan teknologi yang spesifik sehingga mampu mengungkapkan insight dari kumpulan data yang beragam, kompleks dan skala besa (Kampakis, 2016).

Menurut (Agarwal et al., 2017), Big data memiliki karakteristik penting dalam perannya sebagai tools untuk mengelola kompleksitas data yang besar, adalah:

1. Volume mengacu pada semua jenis data yang dapat dihasilkan dari beragam sumber dengan ukuran data yang tumbuh terus menerus.
2. Velocity mengacu pada kecepatan di mana data dibuat, ditransfer dan diproses.
3. Variety merujuk pada berbagai jenis data yang dikelompokkan menjadi structured, semi-structured dan unstructured data.

2.2. Machine Learning

Machine Learning adalah praktik pemrograman komputer untuk belajar dan mengekstraksi pengetahuan dari data. Ini adalah bidang penelitian di persimpangan statistik, kecerdasan buatan, dan ilmu komputer dan juga dikenal sebagai analitik prediktif atau pembelajaran statistik.

Menurut (Han et al., 2012) Pendekatan machine learning memiliki 4 kategori yang umumnya diaplikasikan pada konsep data mining:

1. Supervised learning sering disebut sebagai metode klasifikasi
2. Unsupervised learning merupakan metode pengelompokan (clustering).
3. Semi-supervised learning merupakan teknik machine learning dimana untuk mempelajari model algoritma menggunakan dua jenis data yaitu labeled data dan unlabeled data.
4. Active learning, salah satu pendekatan machine learning dimana user berperan aktif dalam proses pendekatan.

2.3. K-mean clustering

Clustering adalah metode pengelompokkan data, dimana data-data atau objek yang berada dalam 1 *cluster* atau kelompok harus memiliki karakteristik yang mirip satu dengan

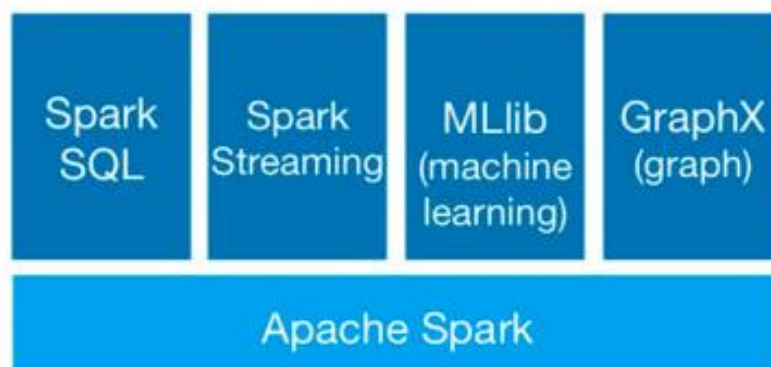
yang lain (Han et al., 2012), sedangkan karakteristik antara 1 *cluster* dengan *cluster* lainnya harus berbeda.

K-Means Clustering bertujuan untuk meminimalisasikan *objective function* yang di *set* dalam proses *clustering*, artinya meminimalkan variasi antar data yang ada dalam 1 *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya. Kelebihan dari metode *K-Means* adalah tidak ada *overlap* data atau data yang menumpuk, 1 data hanya memiliki 1 *cluster*, dapat menganalisis *sampel* dalam jumlah yang lebih besar dengan lebih efisien, kesalahan pada *data outlier* sedikit, hanya ada sedikit perbedaan ukuran jarak yang digunakan, dan *variable* tak relevan atau *variable* yang tidak tepat sedikit jumlahnya. Dasar algoritma *K-Means* adalah sebagai berikut (Han et al., 2012) :

1. Bangkitkan k *centroid* (titik pusat kluster) awal secara acak
2. Hitung jarak yang ada pada setiap data ke masing-masing *centroid* menggunakan Rumus korelasi antar dua objek yaitu *Euclidean Distance*
3. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *Centroid*nya
4. Tentukan posisi *centroid* baru dengan cara melakukan perhitungan nilai rata-rata dari data-data yang ada pada *centroid* yang sama

2.3. Apache Spark

Apache Spark adalah sebuah mesin pengolahan data hadoop memiliki fungsi analisis data yang kompleks dan fungsi machine learning serta algoritma grafik. Apache Spark bisa menjalankan program 100 kali lebih cepat dalam memori dan 10 kali lebih cepat pada disk dibandingkan MapReduce. Apache Spark dapat berjalan beriringan dengan sistem penyimpanan Hadoop dan HDFS. Aplikasi pada Apache Spark dapat dilihat pada gambar 2.1.



(Sumber: <https://spark.apache.org/images/spark-stack.png>)

Gambar 1. Apache Spark

Ada empat komponen utama pada Apache Spark yang dapat digunakan untuk mendukung proses analisa dalam Big Data, adalah:

1. Spark-SQL
komponen di atas Spark yang memperkenalkan abstraksi data baru yang disebut DataFrame, yang menyediakan dukungan untuk data terstruktur dan semi-terstruktur.
2. Spark Streaming
Yang bertujuan untuk memanfaatkan penjadwalan cepat untuk melakukan analisis streaming. Dibutuhkan untuk menganalisis data secara realtime.
3. MLlib (Machine Learning)
Spark MLlib adalah kerangka mesin pembelajaran dan statistik terdistribusi di atas Spark. Dukungan Spark MLlib sebagai algoritma machine learning dan statistic
4. GraphX (graph)
GraphX adalah kerangka pengolahan grafik terdistribusi di atas Apache Spark. Karena didasarkan pada RDDs, yang berubah, grafik yang berubah dan dengan demikian GraphX tidak cocok untuk grafik yang perlu diperbarui, apalagi dengan cara transaksional seperti database grafik.

2.4. Penelitian Terdahulu

Penelitian terdahulu (*previous researches*) ini menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Berikut adalah penelitian terdahulu berupa jurnal terkait dengan penelitian yang dilakukan penulis.

Tabel 1. Penelitian Terdahulu

Nama peneliti	Judul penelitian	Hasil penelitian
Ben Daniel. (Daniel, 2015)	<i>Big Data and analytics in higher education: Opportunities and challenges</i>	Menguraikan sejumlah peluang dan tantangan terkait dengan implementasi Big Data dalam konteks pendidikan tinggi. Penelitian ini menguraikan arah masa depan yang berkaitan dengan pengembangan dan implementasi proyek kelembagaan pada Big Data
Julius Murumba,	<i>Big Data Analytics in Higher Education: A Review</i>	Big Data penting karena menawarkan kepada Universitas peluang untuk sumber daya Teknologi Informasi mereka secara strategis

<p>Elyjoy Micheni. (Murumba & Micheni, 2017)</p>		<p>untuk meningkatkan kualitas pendidikan dan membimbing siswa, perguruan tinggi dan universitas untuk melihat nilai dalam analitik; dan karena itu merekomendasikan agar institusi ini melakukan investasi dalam program analitik dan pada orang-orang untuk memiliki ilmu data yang relevan</p>
<p>Chowdam Sreedhar, Nagulapally Kasiviswanath and Pakanti Chenna Reddy (Sreedhar et al., 2017)</p>	<p><i>Clustering large datasets using K-means modified inter and intra clustering (KM-I2C) in Hadoop</i></p>	<p>Penelitian ini menyajikan dua pendekatan untuk pengelompokan set data besar menggunakan MapReduce. Pendekatan pertama, K-Means Hadoop MapReduce (KM-HMR), berfokus pada implementasi MapReduce dari K-means standar. Pendekatan kedua meningkatkan kualitas cluster untuk menghasilkan cluster dengan jarak intra-cluster maksimum dan jarak antar-cluster minimum untuk dataset besar. Hasil dari pendekatan yang diusulkan menunjukkan peningkatan yang signifikan dalam efisiensi pengelompokan dalam hal waktu pelaksanaan</p>
<p>Prableen Kaura, Manik Sharmab, Mamta Mittal (Kaur et al., 2018)</p>	<p><i>Big Data and Machine Learning Based Secure Healthcare Framework</i></p>	<p>Studi empiris dilakukan untuk menganalisis peran big data dalam industri perawatan kesehatan. Telah diamati bahwa pekerjaan signifikan telah dilakukan dengan menggunakan data besar di sektor perawatan kesehatan. Saat ini, sulit membayangkan cara pembelajaran mesin dan data besar dapat memengaruhi industri perawatan kesehatan. Telah diamati bahwa sebagian besar penulis yang mengimplementasikan penggunaan pembelajaran mesin dan analitik data besar dalam diagnosis penyakit belum memberikan</p>

		bobot yang signifikan pada privasi dan keamanan data
Vinay A dkk. (Vinay et al., 2015)	<i>Cloud Based Big Data Analytics Framework for Face Recognition in Social Networks using Machine Learning</i>	survei menyeluruh tentang konsep Cloud Computing, Big Data, jaringan sosial dan Pembelajaran Mesin dari perspektif kontemporer FR, dan menawarkan kerangka kerja untuk pendekatan FR baru berdasarkan teknik Mesin Pembelajaran Ekstrim untuk melakukan tugas Penandaan Wajah untuk Jaringan Sosial yang beroperasi pada Big Data

BAB III

TUJUAN DAN MANFAAT PENELITIAN

3.1. Tujuan Penelitian

Tujuan Penelitian ini untuk impelementasi big data analytical untuk meningkatkan pengambilan keputusan pihak kampu. Penelitian ini dimulai dengan identifikasi data-data proses berdasarkan learning analitycal, academic dan proses dilingkungan kampus

3.2. Manfaat Penelitin

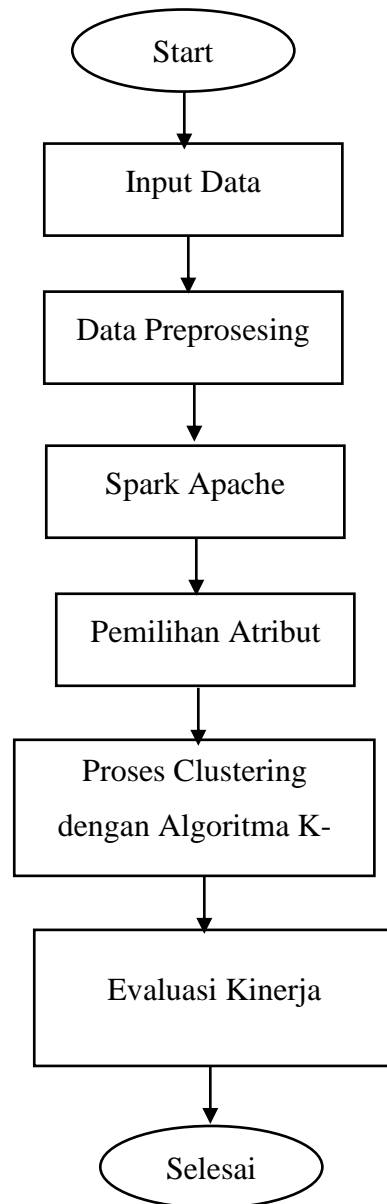
Manfaat penelitian ini kedepannya diharapkan dapat diterapkan dalam bidang pengambilan keputusan pada perguruan tinggi dalam memilih metode analisis data yang besar.

BAB IV METODE PENELITIAN

3.1 Kerangka Pikir

Adapun kerangka pikir dalam penelitian ini adalah dapat dilihat pada gambar 2.

=



Gambar 2. Kerangka Penelitian

3.2 Tahapan Penelitian

Tahapan dalam penelitian ini dapat dilihat pada gambar 3.1.

3.2.1. 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.

3.2.2. 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.

3.2.3. 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:

1. Anaconda 3 (paketan aplikasi yang didalamnya terdapat python, Jupiter dan tool data science lainnya).
2. Spark apache
3. Hadoop
4. Java development kit (JDK)
5. 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)*.

3.2.4 Data selection (pemilihan Attribut)

Data set dengan format *csv* dapat diproses dengan Bahasa pemrograman Python. Maka, rentang data yang akan diolah harus didefinisikan terlebih dahulu. Cara melakukan seleksi rentang data adalah:

1. Impor fungsi `arrange` dari library `numpy`
2. Buat sebuah variabel yang menampung fungsi `arrange`
3. Definisikan secara numerik jumlah baris data yang ingin diseleksi Proses ini akan menghasilkan variabel yang menampung data yang telah diseleksi.

3.2.5. K-mean Clustering

Dalam tahapan ini adalah mengelompokkan data berdasarkan jarak terdekat menggunakan Bahasa pemrograman Python, berikut tahapan k-mean clustering dalam penelitian ini adalah:

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.

3.2.6. 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.

3.3 Metode Pengumpulan data

di unduh dari <https://archive.ics.uci.edu/ml/datasets/Student+Performance> mengenai Penelitian ini menggunakan sumber data atau dataset yang bersifat *public* dapat *Students' Academic Performance*. Beberapa informasi mengenai dataset ini adalah:

1. Judul data adalah *Students' Academic Performance*.
2. Penyumbang data adalah Paulo Cortez, University of Minho, GuimarÃ£es, Portugal, <http://www3.dsi.uminho.pt/pcortez>
3. Dalam penelitian ini dari jumlah 32 atribut diambil 4 yang akan digunakan dalam variable yang akan diolah oleh k-mean clustering
4. Karakteristik data adalah multivariate
5. Jumlah baris data adalah 395

6. Beberapa informasi data dapat dilihat pada tabel 2.

Tabel 2. Informasi Data Student Performance

No	Attribut	Keterangan
1	School	student's school
2	Sex	jenis kelamin mahasiswa
3	Age	Umur Mahasiswa
4	Address	Jenis alamat rumah siswa
5	Famsize	Ukuran Keluarga
6	Pstatus	Status kohabitasi orang tua
7	Medu	Pendidikan Ibu
8	Fedu	Pendidikan Bapak
9	Mjob	Pekerjaan Ibu
10	Fjob	Pekerjaan Bapak
11	Reason	Alasan mengambil sekolah ini
12	Guardian	Wali Mahasiswa
13	Traveltime	Waktu perjalanan pulang dari Kampus
14	StudyTime	Waktu Belajar
15	Failures	jumlah kegagalan kelas sebelumnya
16	Schoolsup	Dukungan pendidikan ekstra
17	Famsup	Dukungan pendidikan keluarga
18	Paid	kelas berbayar ekstra dalam subjek kursus
19	activities	Aktivitas ekstrakurikuler
20	Nursery	menghadiri sekolah pembibitan
21	higr	Ingin mengambil pendidikan tinggi
22	Internet	Internet akses dirumah
23	Romantic	with a romantic relationship
24	Famrel	kualitas hubungan keluarga

25	freetime	Waktu Bebas setelah belajar
26	Gout	pergi bersama teman-teman
27	Dalc	konsumsi alkohol hari kerja
28	Walc	Konsumsi alkohor hari libur
29	health	Status kesehatan sekarang
30	absences	Jumlah absen belajar
31	G1	Grade 1
32	G2	Grade 2
33	G3	Grade3

Untuk melihat data lengkap beserta jumlah record dapat dilihat pada lampiran.

3.3 Metode Analisis

Software yang dipakai adalah browser, python sebagai Bahasa pemograman, untuk mengolah dan menganalisis data menggunakan algoritma K-Mean, dan apache spark untuk mengolah big data.

BAB V

IMPLEMENTASI

Untuk implementasi analisis big data membutuhkan kebutuhan system agar big data bisa digunakan. Beberapa teknologi big data yang digunakan adalah Apache Hadoop sebagai komponen utama dalam manajemen data pada Big Data yang terdiri dari HDFS, Yarn, dan MapReduce. Perancangan big data dimulai dengan instalasi

4.1 Spesifikasi Kebutuhan System

Analisa kebutuhan system digunakan sebagai perangkat lunak dan perangkat keras untuk dapat menjalankan Big Data.

4.1.1 Spesifikasi Perangkat Keras

Spesifikasi perangkat keras yang digunakan dalam penelitian ini dapat dilihat pada tabel 3.

Tabel 3. Spesifikasi Hardware

Prosesor	Intel (T) Core TM i5-3320M 2.60Ghz
RAM	12 GB
VGA	Intel Graphics 4000
Network Adapter	Gigabit Network connection
Hardisk	500GB

4.1.2 Spesifikasi Perangkat Lunak

Spesifikasi perangkat Lunak yang digunakan dalam penelitian ini sesuai tabel 4 berikut :

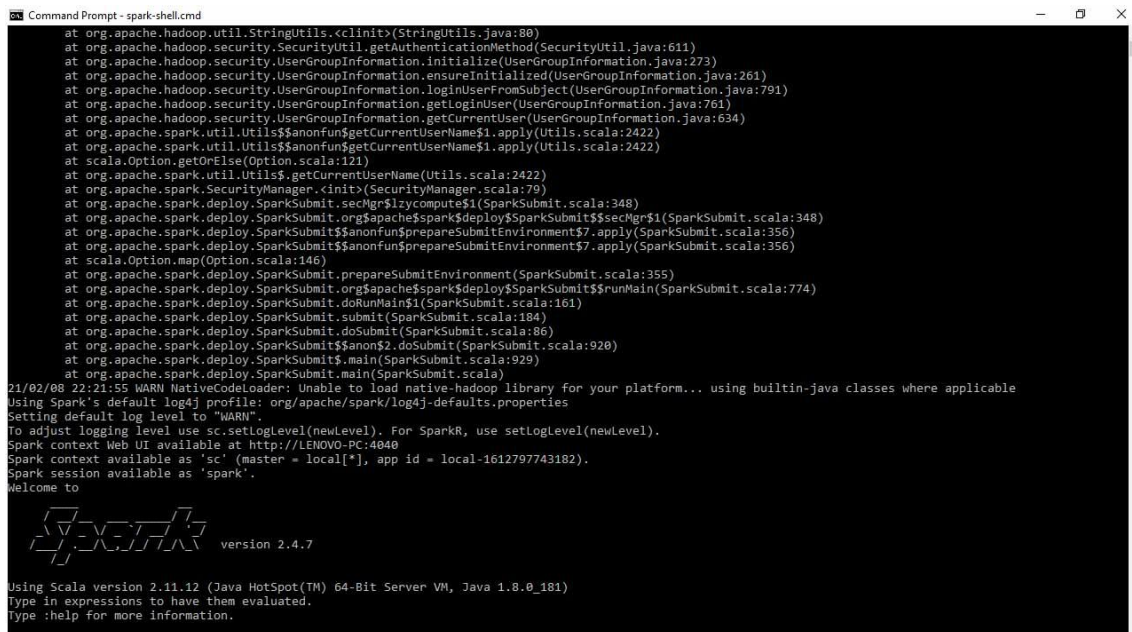
Tabel 4. Spesifikasi Software

Spark apache	Versi 2.70
Sun Java	JDK 8.0
Operating system	Window 10
Hadoop	Versi 2.4
Scala	Versi 2.12

4.2 Perancangan Big Data

4.2.1 Install Spark Apache

Apache spark digunakan untuk sebagai engine untuk memproses big data yang merupakan data berskala besar. Spark apache menyediakan library untuk mengoperasikan python, distribusi skala, dan menampung Hadoop. Spark apache dalam penelitian ini menggunakan versi 2.4.7 digunakan untuk membuat master computer sebagai server dan mengoperasikan akses client menggunakan worker. Berikut tampilan spark apache yang sudah terinstall di komputer.



```
Command Prompt - spark-shell.cmd
at org.apache.hadoop.util.StringUtils.<clinit>(StringUtils.java:80)
at org.apache.hadoop.security.SecurityUtil.getAuthenticationMethod(SecurityUtil.java:611)
at org.apache.hadoop.security.UserGroupInformation.initialize(UserGroupInformation.java:273)
at org.apache.hadoop.security.UserGroupInformation.ensureInitialized(UserGroupInformation.java:261)
at org.apache.hadoop.security.UserGroupInformation.loginUserFromSubject(UserGroupInformation.java:791)
at org.apache.hadoop.security.UserGroupInformation.getLoginUser(UserGroupInformation.java:761)
at org.apache.hadoop.security.UserGroupInformation.getCurrentUser(UserGroupInformation.java:634)
at org.apache.spark.util.Utils$$anonfun$getCurrentUserName$1.apply(Utils.scala:2422)
at org.apache.spark.util.Utils$$anonfun$getCurrentUserName$1.apply(Utils.scala:2422)
at scala.Option.getOrElse(Option.scala:121)
at org.apache.spark.util.Utils$.getCurrentUserName(Utils.scala:2422)
at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
at org.apache.spark.deploy.SparkSubmit$.secMgr$lzycompute$1(SparkSubmit.scala:348)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$secMgr$1(SparkSubmit.scala:348)
at org.apache.spark.deploy.SparkSubmit$$anonfun$prepareSubmitEnvironment$7.apply(SparkSubmit.scala:356)
at org.apache.spark.deploy.SparkSubmit$$anonfun$prepareSubmitEnvironment$7.apply(SparkSubmit.scala:356)
at scala.Option.map(Option.scala:146)
at org.apache.spark.deploy.SparkSubmit.prepareSubmitEnvironment(SparkSubmit.scala:355)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:774)
at org.apache.spark.deploy.SparkSubmit.doRunMain$1(SparkSubmit.scala:161)
at org.apache.spark.deploy.SparkSubmit.submit(SparkSubmit.scala:184)
at org.apache.spark.deploy.SparkSubmit.doSubmit(SparkSubmit.scala:86)
at org.apache.spark.deploy.SparkSubmit$$anonfun$doSubmit(SparkSubmit.scala:920)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:929)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
21/02/08 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 Web UI available at http://LENOVO-PC:4040
Spark context available as 'sc' (master = local[*], app id = local-1612797743182).
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_181)
Type in expressions to have them evaluated.
Type :help for more information.
```

Gambar 3. Tampilan Spark Apache

4.2.2 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 adalah sebagai berikut:

1. Install Java Development Kit (JDK)
2. Buat user dan group untuk Hadoop
3. Konfigurasi firewall dinonaktifkan

4. Instalasi Hadoop dan konfigurasi Hadoop Environment pada node master dan slave Untuk menjalankan

4.2.3 Konfigurasi Environment

Pada tahapan ini adalah setting environment variable pada my computer agar big data dapat digunakan. Setting environment dapat dilihat pada tabel 5.

Tabel 5. 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	
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

Tabel 5 adalah setting environment untuk menjalankan python dengan instalasi anaconda 3 yang berisikan Jupiter, notebook dan python 3.

4.2.4 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.

Tampilan akses ke master slave dapat dilihat pada gambar 4.

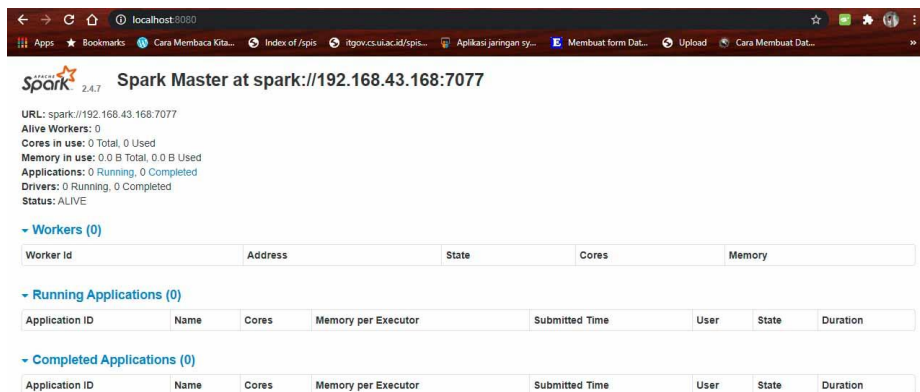
```

Anaconda Prompt (python) - spark-class org.apache.spark.deploy.master.Master
(base) C:\Users\LENOVO>cd C:\spark\spark\bin
(base) C:\spark\spark\bin>spark-class org.apache.spark.deploy.master.Master
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/02/08 23:21:12 INFO Master: Started daemon with process name: 10412@LENOVO-PC
21/02/08 23:21:13 ERROR Shell: Failed to locate the winutils binary in the hadoop binary path
java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.
    at org.apache.hadoop.util.Shell.getQualifiedBinPath(Shell.java:379)
    at org.apache.hadoop.util.Shell.getWinUtilsPath(Shell.java:394)
    at org.apache.hadoop.util.Shell.<clinit>(Shell.java:387)
    at org.apache.hadoop.util.StringUtils.<clinit>(StringUtils.java:80)
    at org.apache.hadoop.security.SecurityUtil.getAuthenticationMethod(SecurityUtil.java:611)
    at org.apache.hadoop.security.UserGroupInformation.initialize(UserGroupInformation.java:273)
    at org.apache.hadoop.security.UserGroupInformation.ensureInitialized(UserGroupInformation.java:261)
    at org.apache.hadoop.security.UserGroupInformation.loginUserFromSubject(UserGroupInformation.java:791)
    at org.apache.hadoop.security.UserGroupInformation.getLoginUser(UserGroupInformation.java:761)
    at org.apache.hadoop.security.UserGroupInformation.getCurrentUser(UserGroupInformation.java:634)
    at org.apache.spark.util.Utils$.anonFun$getCurrentUserName$1.apply(Utils.scala:2422)
    at org.apache.spark.util.Utils$.anonFun$getCurrentUserName$1.apply(Utils.scala:2422)
    at scala.Option.getOrElse(Option.scala:121)
    at org.apache.spark.util.Utils$.getCurrentUserName(Utils.scala:2422)
    at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
    at org.apache.spark.deploy.master.Master.startRpcEnvAndEndpoint(Master.scala:1073)
    at org.apache.spark.deploy.master.Master$.main(Master.scala:1058)
    at org.apache.spark.deploy.master.Master.main(Master.scala)
21/02/08 23:21:13 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
21/02/08 23:21:13 INFO SecurityManager: Changing view acls to: LENOVO
21/02/08 23:21:13 INFO SecurityManager: Changing modify acls to: LENOVO
21/02/08 23:21:13 INFO SecurityManager: Changing view acls groups to:
21/02/08 23:21:13 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(LENOVO); groups with view permissions: Set(); users with modify permissions: Set(LENOVO); groups with modify permissions: Set()
21/02/08 23:21:14 INFO Utils: Successfully started service 'sparkMaster' on port 7077.
21/02/08 23:21:14 INFO Master: Starting Spark master at spark://192.168.43.168:7077
21/02/08 23:21:14 INFO Master: Running Spark version 2.4.7
21/02/08 23:21:15 INFO Utils: Successfully started service 'MasterUI' on port 8080.
21/02/08 23:21:15 INFO MasterWebUI: Bound MasterWebUI to 0.0.0.0, and started at http://LENOVO-PC:8080
21/02/08 23:21:15 INFO Master: I have been elected leader! New state: ALIVE

```

Gambar 4. Tampilan koneksi Master Slave

Gambar 5 adalah tampilan koneksi master slave berhasil dieksekusi. Untuk akses spark master yang digunakan untuk mengelola worker sebagai client atau aplikasi running yang dilakukan oleh computer slave. Untuk masuk ke spark master menggunakan localhost:8080. Tampilan spark apaceh GUI dapat dilihat pada gambar 4.3.



Gambar 5 Tampilan spark master

2. Koneksi Komputer Worker

Computer 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. Tampilan koneksi dari worker ke master dapat dilihat pada gambar 6 dan 7

```

(base) C:\Users\LENOVO>cd C:\spark\spark\bin
(base) C:\spark\spark\bin>spark-class org.apache.spark.deploy.worker.Worker spark://192.168.43.168:7077
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/02/08 23:38:32 INFO Worker: Started daemon with process name: 5780@LENOVO-PC
21/02/08 23:38:33 ERROR Shell: Failed to locate the winutils binary in the hadoop binary path
java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.
    at org.apache.hadoop.util.Shell.getQualifiedBinPath(Shell.java:379)
    at org.apache.hadoop.util.Shell.getWinutilsPath(Shell.java:394)
    at org.apache.hadoop.util.Shell.<clinit>(Shell.java:367)
    at org.apache.hadoop.util.StringUtil.<clinit>(StringUtil.java:88)
    at org.apache.hadoop.security.SecurityUtil.getAuthenticationMethod(SecurityUtil.java:611)
    at org.apache.hadoop.security.UserGroupInformation.initialize(UserGroupInformation.java:273)
    at org.apache.hadoop.security.UserGroupInformation.ensureInitialized(UserGroupInformation.java:261)
    at org.apache.hadoop.security.UserGroupInformation.loginUserFromSubject(UserGroupInformation.java:791)
    at org.apache.hadoop.security.UserGroupInformation.getLoginUser(UserGroupInformation.java:761)
    at org.apache.hadoop.security.UserGroupInformation.getCurrentUser(UserGroupInformation.java:634)
    at org.apache.spark.util.Utils$.anonfun$getCurrentUserName$1.apply(Utils.scala:2422)
    at scala.Option.getOrElse(Option.scala:121)
    at org.apache.spark.util.Utils$.getCurrentUserName(Utils.scala:2422)
    at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
    at org.apache.spark.deploy.worker.Worker$.startUpEnvAndBootstrap(Worker.scala:808)
    at org.apache.spark.deploy.worker.Worker$.main(Worker.scala:779)
    at org.apache.spark.deploy.worker.Worker.main(Worker.scala)
21/02/08 23:38:33 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
21/02/08 23:38:33 INFO SecurityManager: Changing view acls to: LENOVO
21/02/08 23:38:33 INFO SecurityManager: Changing modify acls to: LENOVO
21/02/08 23:38:33 INFO SecurityManager: Changing view acls groups to:
21/02/08 23:38:33 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(LENOVO); groups with view permissions: Set(); users with modify permissions: Set(LENOVO); groups with modify permissions: Set()
21/02/08 23:38:34 INFO Utils: Successfully started service 'sparkworker' on port 51114.
21/02/08 23:38:35 INFO Worker: Starting Spark worker 192.168.43.168:5114 with 4 cores, 10.7 GB RAM
21/02/08 23:38:35 INFO Worker: Running Spark version 2.4.7
21/02/08 23:38:35 INFO Worker: Spark home: C:\spark\spark\bin\..
21/02/08 23:38:35 INFO Utils: Successfully started service 'workerUI' on port 8081.
21/02/08 23:38:35 INFO WorkerWebUI: Bound WorkerWebUI to 8.0.8.0, and started at http://LENOVO-PC:8081
21/02/08 23:38:35 INFO Worker: Connecting to master 192.168.43.168:7077...
21/02/08 23:38:35 INFO TransportClientFactory: Successfully created connection to 192.168.43.168:7077 after 114 ms (0 ms spent in bootstraps)
21/02/08 23:38:36 INFO Worker: Successfully registered with master spark://192.168.43.168:7077
  
```

Gambar 6. Tampilan koneksi worker ke master slave

Gambar 6 Ketika worker sudah terkoneksi ke master slave, maka id worker akan tampil di spark apace master slave.

Spark Master at spark://192.168.43.168:7077

URL: spark://192.168.43.168:7077
 Alive Workers: 1
 Cores in use: 4 Total, 0 Used
 Memory in use: 10.7 GB Total, 0.0 B Used
 Applications: 0 Running, 0 Completed
 Drivers: 0 Running, 0 Completed
 Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory
worker-20210208233834-192.168.43.168-51114	192.168.43.168:51114	ALIVE	4 (0 Used)	10.7 GB (0.0 B Used)

Running Applications (0)

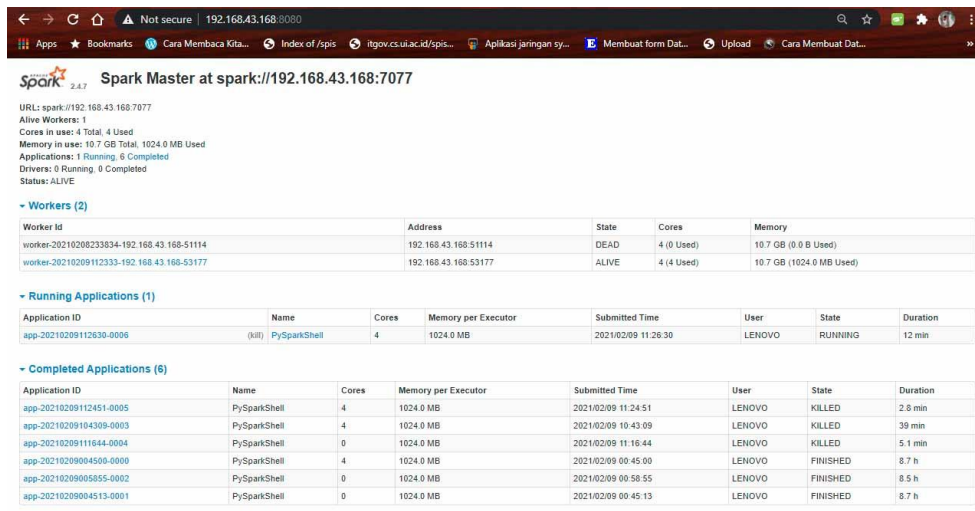
Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

Gambar 7. Tampilan spark master deteksi worker

Tampilan ketika ada computer worker dalam menjalankan aktivitas dapat dilihat pada gambar 8.



Gambar 8 Tampilan Running Aktivitas

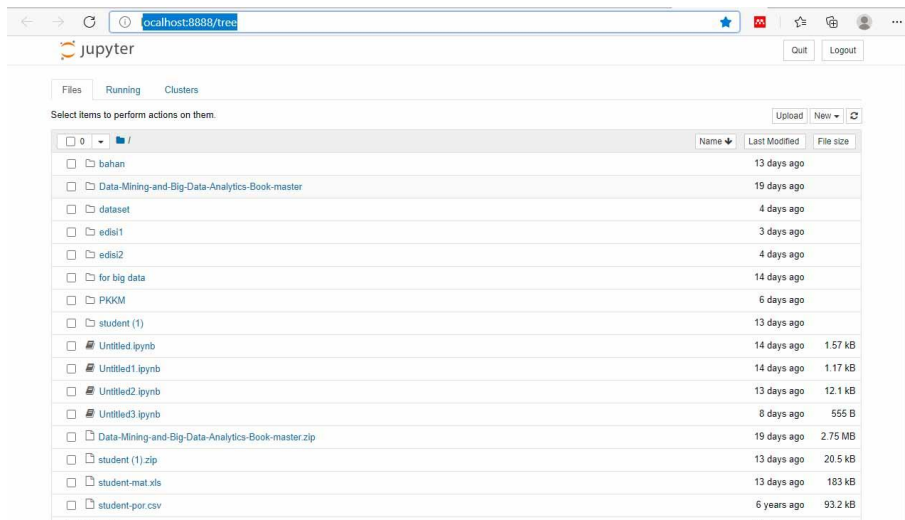
3. Menjalankan pyspark

Setelah master slave dan worker dijalankan maka pyspark dapat dijalankan. Memanggil pyspark adalah untuk masuk ke directory dimana data-data atau menjalankan python dilakukan. Gambar 9 tahapan untuk menjalankan pyspark dan gambar 10 tampilan ke localhost:8888/tree setelah pyspark telah sukses di koneksi.

```
Anaconda Prompt (python) - pyspark --master spark://192.168.43.168:7077
(base) C:\Users\LENOVO>E:
(base) E:\>CD RISET INTERNAL
(base) E:\riset internal>pyspark --master spark://192.168.43.168:7077
[I 23:57:11.290 NotebookApp] JupyterLab extension loaded from E:\python\lib\site-packages\jupyterlab
[I 23:57:11.290 NotebookApp] JupyterLab application directory is E:\python\share\jupyter\lab
[I 23:57:11.308 NotebookApp] Serving notebooks from local directory: E:\riset internal
[I 23:57:11.309 NotebookApp] The Jupyter Notebook is running at:
[I 23:57:11.310 NotebookApp] http://localhost:8888/?token=edc3fe23f4c38b288fabf06c4f779b6e898786473ecf160e
[I 23:57:11.310 NotebookApp] or http://127.0.0.1:8888/?token=edc3fe23f4c38b288fabf06c4f779b6e898786473ecf160e
[I 23:57:11.310 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 23:57:11.419 NotebookApp]

To access the notebook, open this file in a browser:
file:///C:/Users/LENOVO/AppData/Roaming/jupyter/runtime/nbserver-244-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=edc3fe23f4c38b288fabf06c4f779b6e898786473ecf160e
or http://127.0.0.1:8888/?token=edc3fe23f4c38b288fabf06c4f779b6e898786473ecf160e
```

Gambar 9. Menjalankan Pyspark

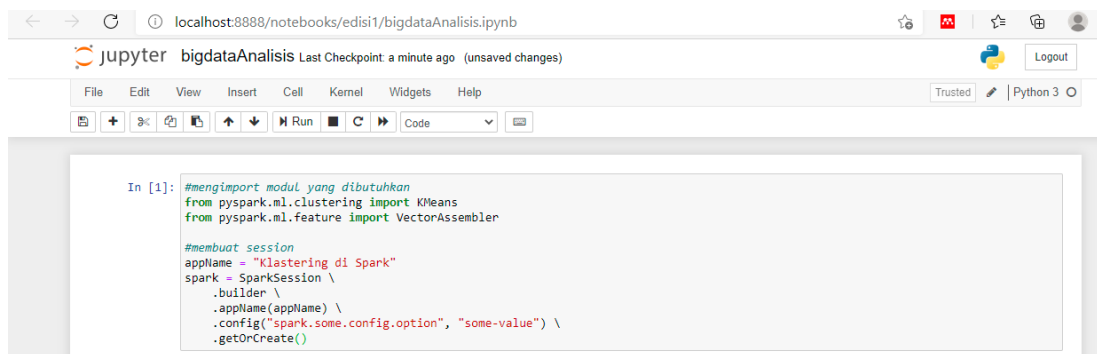


Gambar 10. Directory Jupyter

4.3 Implementasi Kmean Clustering menggunakan Spark Apache

4.3.1 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 Jupyter. Berikut listing program untuk persiapan input data dalam dilihat pada Gambar 11.



Gambar 11. list program untuk import library

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 12.


```
In [6]: #memuat data dari file ke DataFrame dengan infer skema
customers = spark.read.csv(
'dataset/student_GPA.csv', inferSchema=True, header=True)
customers.show(396)
```

absences	G1	G2	G3
6	5	6	6
4	5	5	6
10	7	8	10
2	15	14	15
4	6	10	10
10	15	15	15
0	12	12	11
6	6	5	6
0	16	18	19
0	14	15	15
0	10	8	9
4	10	12	12
2	14	14	14
2	10	10	11
0	14	16	16
4	14	14	14

Gambar 12. list program untuk import library

4.3.2 Data Training

Tahap selanjutnya adalah menyiapkan data training sebelum memproses K-mean clusteringnya. Dimana dalam tahap ini sebelum dimasukan ke model kmean data dirubah kedalam bentuk vector assembler. Berikut listing program tahap training menggunakan python dapat dilihat pada gambar 13.

```
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
[6.0,5.0,6.0,6.0]
[4.0,5.0,5.0,6.0]
[10.0,7.0,8.0,10.0]
[2.0,15.0,14.0,15.0]
[4.0,6.0,10.0,10.0]
[10.0,15.0,15.0,15.0]
[0.0,12.0,12.0,11.0]
[6.0,6.0,5.0,6.0]
[0.0,16.0,18.0,19.0]
[0.0,14.0,15.0,15.0]

only showing top 10 rows

Gambar 13 list program untuk Data Training

4.3.3 Membuat Model K-mean Clustering

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 14.

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

Gambar 14 Membangun Model K-mean

4.3.4 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 pemrograman python, ditampilkan pada gambar 15.

```
In [4]: 1 #mendefinisikan algoritma klustering
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 [29]: 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]
[3.34013605 9.24489796 9.6122449 9.7414966 ]
[ 3.04724409 14.48818898 14.58267717 14.83464567]
[52.6 10.2 10. 9.4]
[15.19444444 10.23611111 9.80555556 9.95833333]
```

Gambar 15. Listing Program Untuk Menentukan Centroid

4.3.5 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 16.

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

Gambar 16. Listing Program Untuk Prediksi

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 17.

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

Gambar 17. Hasil prediksi cluster

Hasil prediksi dari setiap record yang telah dilakukan proses clustering dengan k-mean dapat dilihat dari gambar 18.

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

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

Gambar 18. Hasil Prediksi Tiap Record

4.4 Hasil Analisis implementasi Big Data

Dalam penelitian ini menghasilkan pengelompokan data dari dataset student performace 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.

BAB VI

KESIMPULAN

5.1 Kesimpulan

Dari hasil penelitian implementasi analisis big data pada perguruan tinggi menggunakan machine learning disimpulkan sebagai berikut:

1. 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.
2. Dalam penelitian Analisis big data dioperasikan menggunakan apache spark selanjutnya untuk proses pengelompokan data menggunakan algoritma k-mean clustering bagian dari algoritma machine learning.
3. Hasil kinerja mahasiswa didapat dari 4 variable yaitu nilai abses, 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.

5.2 Saran

Saran yang selanjutnya bisa dilanjutkan dalam penelitian penerapan big data untuk perguruan tinggi menggunakan pendekatan machine learning adalah:

1. Sebaiknya kriteria learning analytical, academic analytical, dan staff analytical harus diolah dan dilakukan analisis bersama-sama.
2. Untuk hasil kinerja big data lebih baik, sebaiknya menggunakan computer server atau *cloud computing*
3. Sebaiknya menggunakan dataset yang langsung diambil dari system informasi akademik tiap perguruan tinggi untuk mendapatkan nilai maksimal dalam penelitian.

REFERENSI

- Agarwal, B., Ravikumar, A., & Saha, S. (2017). A novel approach to big data veracity using crowdsourcing techniques and Bayesian predictors. *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, 1020–1023. <https://doi.org/10.1109/ICMLA.2016.25>
- Asniar. (2015). *Penggunaan Big Data Analytic di Perguruan Tinggi*. June, 1–5. <https://doi.org/10.13140/RG.2.1.4581.9046>
- B.Tulasi. (2013). Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 68(14), 1–4. https://doi.org/10.1007/978-3-319-06520-5_1
- 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>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3ed ed.). Morgan Kauffman.
- 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>
- Kampakis, S. (2016). *Predictive modelling of football injuries*. April. <http://arxiv.org/abs/1609.07480>
- 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. 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>
- Sreedhar, C., Kasiviswanath, N., & Chenna Reddy, P. (2017). Clustering large datasets using

K-means modified inter and intra clustering (KM-I2C) in Hadoop. *Journal of Big Data*, 4(1). <https://doi.org/10.1186/s40537-017-0087-2>

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>

Zhang, Z., Li, J., & Yang, X. (2020). Data Aggregation in Heterogeneous Wireless Sensor Networks by Using Local Tree Reconstruction Algorithm. *Complexity*, 2020. <https://doi.org/10.1155/2020/3594263>

LAMPIRAN-LAMPIRAN

Lampiran 1. Dukungan Sarana dan Prasarana Penelitian

Sarana dan prasarana penelitian yang digunakan dikelompokkan menjadi 3, yaitu hardware, data dan software. Uraianya adalah sebagai berikut :

1. Hardware

Spesifikasi hardware yang digunakan untuk penelitian ini adalah komputer dengan spesifikasi :

Prosesor : Intel (T) Core TM i5-3320M 2.60Ghz

RAM : 12 GB

VGA : Intel Graphics 4000

Network Adapter : Gigabit Network connection

Hardisk : 500GB

2. Data

Data yang digunakan merupakan Dataset Student Performnace yang dapat di download pada link

<http://www3.dsi.uminho.pt/pcortez>

3. Software

Spesifikasi Software yang digunakan untuk penelitian ini adalah :

Spark apache : Versi 2.70

Sun Java : JDK 8.0

Operating system : Window 10

Hadoop : Versi 2.4

Scala : Versi 2.12

Lampiran 2. Susunan organisasi tim peneliti dan pembagian tugas

NO	NAMA	NIDN	BIDANG ILMU	URAIAN TUGAS
1.	Tri Dharma Putra, S. ST., M.Sc (Ketua)	0302117101	Informatika	<ol style="list-style-type: none"> 1. Bertanggung jawab atas hasil penelitian. 2. Memastikan setiap anggota bekerja sesuai tugasnya. 3. Mengarahkan pembuatan e-questioner sebagai instrumen pengumpulan data dan mendistribusikannya. 4. Bertanggung jawab pada proses penelitian. 5. Memeriksa dan membantu proses penilaian dari awal sampai selesai. 6. Melakukan perhitungan manual pada setiap metode 7. Memastikan luaran wajib tercapai 8. Membantu pembuatan aplikasi
2.	Rakhmat Purnomo, S.Pd., S.Kom., M.Kom (Anggota 1)	0322108201	Informatika	<ol style="list-style-type: none"> 1. Membantu tugas ketua dalam penelitian 2. Membantu menyelesaikan laporan hingga tuntas 3. Fokus pada analisa dan perancangan sistem 4. Membantu pendistribusia e-questioner sebagai instrumen pengumpulan data 5. Membuat Program.
3.	Wowon Priatna, ST, M.TI (Anggota 2)	0429118007	Informatika	<ol style="list-style-type: none"> 1. Membantu ketua dalam perhitungan manual. 2. Membantu dalam pembuatan aplikasi dengan Python. 3. Membantu dalam pengujian sistem

Lampiran 3. Biodata Ketua dan Anggota

Ketua Peneliti

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Tri Dharma Putra, ST., M.Sc
2	Jenis Kelamin	AA
3	Jabatan Fungsional	Asisten Ahli
4	NIP/NIK/Identitas lainnya	3275030211710022
5	NIDN	0302117101
6	E-mail	tri.dharma.putra@dsn.ubharajaya.ac.id
7	Nomor Telepon/HP	08170707785
8	Alamat Kantor	Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Kota Bekasi, Jawa Barat 17121, Indonesia
9	Nomor Telepon/Faks	(021) 88955882
10	Mata Kuliah yg Diampu	1. Pemrograman Berorientasi Objek 2. Sistem Operasi 3. Pemrograman Terstruktur

B. Publikasi Jurnal

No	Tahun	Judul Penelitian	Nama Jurnal
1	2018	Perbandingan Tahapan Algoritma Apriori Klasik dengan Apriori Termodifikasi	Jurnal Kajian Ilmiah Universitas Bhayangkara Jaya
2	2018	Meningkatkan Kualitas Siswa SMA Muhammadiyah 1 Jakarta dengan Pemanfaatan Aplikasi OpenOffice sebagai Media	Jurnal Abdimas Universitas
3	2020	Pemberdayaan Sumber Daya Desa Sukawijaya melalui Teknologi Informasi	Jurnal Sains Teknologi dalam Pemberdayaan Masyarakat
4	2020	Manfaat dan Metode Aplikasi Sensus Penduduk Online di Desa Sukabakti	Jurnal Pengabdian Masyarakat Universitas Islam Nusantara (UNINUS)
5	2020	Analysis of Preemptive Shortest Job First (SJF) Algorithm in CPU Scheduling	International Journal of Advanced Research in Computer and Communication Engineering
6	2020	Analisis Keranjang Belanja dengan Algoritma Apriori Klasik pada Data Mining	Jurnal Kajian Ilmiah Universitas Bhayangkara Jaya

Ketua Peneliti

Tri Dharma Putra, ST., M.Sc

Anggota Peneliti 1

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Rakhmat Purnomo, S.Pd., S.Kom., M.Kom
2	Jenis Kelamin	L
3	Jabatan Fungsional	Asisten Ahli
4	NIP/NIK/Identitas lainnya	3275062210820026
5	NIDN	0322108201
6	E-mail	rakhmat.purnomo@dsn.ubharajaya.ac.id
7	Nomor Telepon/HP	081383013873
8	Alamat Kantor	Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Kota Bekasi Jawa Barat 17121 Indonesia
9	Nomor Telepon/Faks	(021) 88955882
10.	Mata Kuliah yg Diampu	1. Jaringan Komputer 2. Manajemen Proyek 3. Sistem Operasi

B. Riwayat Penelitian

NO.	Tahun Jabatan	Judul	Skema Sumber Dana
1	2019 Ketua Pengusul	Perbandingan Fuzzy Logic dan Analytic Hierarchy Process untuk Menilai Kinerja Dosen di Universitas Bhayangkara Jakarta Raya	PDP/Dosen Pemula DIKTI
2	2020 Anggota Pengusul	Implementasi Algoritma Genetika dalam Optimasi Penjadwalan Mata Kuliah Berdasarkan Team-Teaching pada Program Studi Teknik Informatika UBJ	PDP/Dosen pemula DIKTI

C. Riwayat Publikasi

No	Tahun/Jenis Publikasi/Status	Nama Judul dan Jurnal
	Tahun Publikasi : 2017 Status Penulis : ke 1	Judul artikel : Penerapan Greedy Forward Selection dan Bagging pada Logistic Regression untuk Prediksi Cacat Perangkat Lunak Nama Jurnal : Jurnal Kajian Ilmiah Vol 17, No 2 (2017)
	2019 Penulis ke 1 Jurnal Nasional tidak terakreditasi (mempunyai ISSN)	Perbandingan Logika Fuzzy dan Analytic Hierarchy Process untuk Menilai Kinerja Dosen Nama Jurnal : Jurnal ESIT ISSN: 1979-1909 Vol. 14 No. 3 Oktober 2019 Penerbit : STMIK Eresha
	2019 Penulis ke 2 Jurnal Nasional	Pengendali Robot Tempat Sampah Menggunakan Smartphone Berbasis Boarduino Nama Jurnal : Jurnal Jaring SainTek

No	Tahun/Jenis Publikasi/Status	Nama Judul dan Jurnal
	tidak terakreditasi (mempunyai ISSN)	ISSN : 2656-9485 Vol. 1 No. 2 Oktober 2019
	2019 Penulis ke 1 Jurnal Nasional tidak terakreditasi (mempunyai ISSN)	Analisis QoS dengan Virtual Tenant Network pada Software Define Networking Nama Jurnal : Jurnal Teknologi Informatika dan Komputer ISSN: 2622-8475 Vol 5, No 2 September 2019
	2019 Penulis ke 2 Jurnal Nasional tidak terakreditasi (mempunyai ISSN)	Analisis Luas Lahan Kelapa Sawit Menggunakan Metode FuzzyAHP Terhadap Resiko Pelaku Usaha Jurnal Jaring SainTek ISSN : 2656-9485 Vol 1 No 1 April 2019
	2020 Penulis ke 1 Jurnal Nasional Terakreditasi	Implementasi FingerPrint dan Short Message Service Gateway pada Sistem Presensi Jurnal Kajian Ilmiah (JKI) ISSN :1410-9794, e-ISSN : 2597-792X Vol. 20 No. 1 (Januari 2020), hal 93 – 102 Terakreditasi SINTA 5 sesuai SK RISTEKDIKTI Nomor. 3/E/KPT/2019
	2020 Penulis ke 2 Jurnal Nasional Terakreditasi	Implementasi Fuzzy Inference System Metode Sugeno Pada Aplikasi Penilaian Kinerja Dosen Jurnal Techno.COM: Jurnal Teknologi Informasi eISSN : 2356-2579 Vol 19 No. 3, Agustus 2020 Terakreditasi SINTA 3

Anggota Peneliti 1

Rakhmat Purnomo, S.Pd., S.Kom.,

M.Kom.

Anggota Peneliti 2

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Wowon Priatna
2	Jenis Kelamin	L
3	Jabatan Fungsional	Asisten Ahli
4	NIP/NIK/Identitas lainnya	3175032911800014
5	NIDN	0322108201
6	E-mail	Wowon.priatna@dsn.ubharajaya.ac.id
7	Nomor Telepon/HP	081932035255
8	Alamat Kantor	Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Kota Bekasi Jawa Barat 17121 Indonesia
9	Nomor Telepon/Faks	(021) 88955882
10.	Mata Kuliah yg Diampu	1. Machine Learning 2. Pengantar Rekayasa Perangkat lunak 3. Pemograman Berorientasi Object

A. Pengalaman Penelitian dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber*	Jml (Juta Rp)
1	Februari 2019	Perbandingan Fuzzy logic dan Analytic Hierarchy Process untuk Menilai Kinerja Dosen di Universitas Bhayangkara Jakarta Raya	Dikti	19 Juta
2	Februari 2018	Optimasi Penjadwalan Mata Kuliah Di Prodi Teknik Informatika STT Pelita Bangsa Dengan Menggunakan Algoritma Genetika	Dikti	20 Juta

B. Publikasi Jurnal

No.	Tahun	Judul Penelitian	Nama Jurnal
1	Agustus 2020	Implementasi Fuzzy Inference System Metode Sugeno Pada Aplikasi Penilaian Kinerja Dosen	Techno.com Akreditasi Shinta 3
2	Juni 2020	Peningkatan Lingkungan Bersih dan Sehat pada Desa Kedung Jaya Kecamatan Babelan Bekasi	Jurnal Sains Teknologi dalam Pemberdayaan Masyarakat
3	Desember 2019	Perancangan Sistem Pendukung Keputusan Untuk Pemilihan Lokasi Dalam Perluasan Usaha Kafe menggunakan Analytical Hierarchy Process	Jurnal Resti Akreditasi Shinta 2
4	Oktober 2019	Perbandingan Logika Fuzzy dan Analytic Hierarchy Process untuk Menilai Kinerja Dosen	Jurnal Esit

5	Oktober 2019	A Implementasi Term Frequency–Inverse Document Frequency (Tf-Idf) Dan Vector Space Model (Vsm) Untuk Pencarian Berita Bahasa Indonesia	Jurnal Pelita Bangsa
6	Juli 2019	Sistem Pendukung Keputusan Untuk Penentuan Dosen Favorit Menggunakan Simple Additive Weighting (SAW)	Jurasik
7	Oktober 2018	Implementasi Algoritma Genetika Untuk Optimasi Penjadwalan Mata Kuliah	Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)
8	Oktober 2018	Optimasi Penjadwalan Mata Kuliah Menggunakan Algoritma Genetika Pada Prodi Teknik Informatika Stt Pelita Bangsa	Jurnal conference SAINS DAN TEKNOLOG
9	23 Oktober 2013	Pengaruh Kematangan, Kinerja Dan Pemanfaatan Teknologi Informasi Terhadap Implementasi Si Di SMK Negeri Jakarta Timur Dengan Model Cobit Framework	Jurnal Sistem Informasi (Journal of Information System) Akreditasi Shinta 2

Anggota Peneliti 2

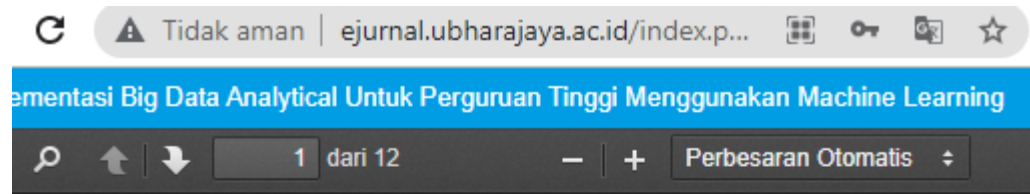
Wowon Priatna, ST., M.TI.

Lampiran 4. Surat Pernyataan Ketua Peneliti

Lampiran 5. Rincian Anggaran

Lampiran 6. Bukti Publikasi (Terbit)

<http://ejurnal.ubharajaya.ac.id/index.php/jiforty/article/view/633/525>



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>

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

Lampiran 7. Log Book Penelitian

Berikut adalah Log-Book Penelitian

NO	TGL	KEGIATAN
1	5 Sept 2020	Tim menyusun Proposal Penelitian
2	16 Sept 2020	Pengajuan usulan proposal
3	19 Sept 2020	Pelaksanaan Review Proposal
4	21 Sept 2020	Tanda tangan Kontrak Penelitian
5	29 Sept 2020	Pencairan 50% dana penelitian
6	1 Okt – 30 des 2020	Pelaksanaan Penelitian
7	5 April 2021	Submit Artikel hasil penelitian (Status In-Review)