

PROSIDING



Seminar Nasional Matematika dan Pembelajarannya (SNMP) 2020

**“Peran Matematika dan Pendidikan Matematika
dalam Menyongsong Era Disrupsi”**



27 Oktober 2020

Diterbitkan oleh:
FMIPA Universitas Negeri Malang
ISBN: 978-623-96664-0-8

Website:
<http://matematika.fmipa.um.ac.id/seminar/2020>

Email:
seminar.matematika.fmipa@um.ac.id

PROSIDING

SEMINAR NASIONAL MATEMATIKA DAN PEMBELAJARANNYA (SNMP) 2020

Tema:

Peran Matematika dan Pendidikan Matematika dalam
Menyongsong Era Disrupsi

Malang, 27 Oktober 2020

Penerbit

FMIPA Universitas Negeri Malang

PROSIDING

SEMINAR NASIONAL MATEMATIKA DAN PEMBELAJARANNYA 2020

“Peran Matematika dan Pendidikan Matematika dalam Menyongsong Era Disrupsi”

Malang, 27 Oktober 2020

Tim Editor:

Mochammad Hafiizh, S.Pd., M.Si., Ph.D.
Vita Kusumasari, S.Si., M.Pd., Ph.D.
Dr. Desi Rahmadani, S.Si., M.Si.
Latifah Mustofa Lestyanto, S.Si., M.Pd.
Mohammad Agung, S.Pd., M.Sc.
Kridha Pusawidjayanti, S.Si., M.Si.
Denis Eka Cahyani, S.Kom., M.Kom.
Asmianto, S.Si., M.Si.
Azizah, S.Si., M.Si.

ISBN : 978-623-96664-0-8

Perpustakaan Nasional: Katalog dalam Terbitan (KDT)

Hak cipta dilindungi undang-undang. Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi buku ke dalam bentuk apapun, secara elektronis maupun mekanis, termasuk fotokopi atau merekam dengan teknik apapun tanpa izin tertulis dari penerbit.

ISBN 978-623-96664-0-8

Penerbit: FMIPA Universitas Negeri Malang
Jl. Semarang 5 Malang
Gedung O1 UM



PROSIDING

SEMINAR NASIONAL MATEMATIKA DAN PEMBELAJARANNYA 2020

STEERING COMMITTEE

Mochammad Hafizh, S.Pd., M.Si., Ph.D
Azizah, S.Pd., M.Si
Kridha Pusawidjayanti, S.Si., M.Si
Dr. Susiswo, M.Si
Dr. Erry Hidayanto, M.Si
Dr. Abd. Qohar, M.T
Mohammad Agung, S.Pd, M.Sc
Latifah Mustofa Lestyanto, S.Si, M.Pd
Vita Kusumasari, S.Si., M.Pd., Ph.D
Dr. Desi Rahmadani, S.Si., M.Si

REVIEWER

Prof. Purwanto, Ph.D.
Prof. Dr. Toto Nusantara, M.Si
Dahliatul Hasanah, S.Si, M.Mat.Sc.
Darmawan Satyananda, S.T, M.T
Dr. Abd. Qohar, M.T
Dr. Desi Rahmadani, S.Si., M.Si.
Dr. Erry Hidayanto, M.Si
Dr. Hery Susanto, M.Si
Dr. I Nengah Parta, S.Pd, M.Si
Dr. Ir. Hendro Permadi, M.Si
Dr. Makbul Muksar, S.Pd, M.Si
Dr.rer.nat. I Made Sulandra, M.Si
Dr. Rustanto Rahardi, M.Si

Dr. Subanji, M.Si
Dr. Sukoriyanto, M.Si
Dr. Swasono Rahardjo, S.Pd, M.Si
Dra. Santi Irawati, M.Si, Ph.D.
Dra. Sapti Wahyuningsih, M.Si
Drs. Tjang Daniel C, M.Si, Ph.D.
Indriati Nurul Hidayah, S.Pd, M.Si
Lucky Tri Oktoviana, S.Si, M.Kom
Mochammad Hafizh, S.Pd, M.Si, Ph.D
Mohamad Yasin, S.Kom, M.Kom
Nur Atikah, S.Si, M.Si
Trianingsih Eni Lestari, S.Si, M.Si
Vita Kusumasari, S.Si, M.Pd, Ph.D

Penerbit: FMIPA Universitas Negeri Malang
Jl. Semarang 5 Malang
Gedung O1 UM

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT, atas limpahan rahmat dan hidayah-Nya sehingga FMIPA Universitas Negeri Malang dapat menyelesaikan Prosiding Seminar Nasional Matematika dan Pembelajarannya 2020 yang mengusung tema “Peran Matematika dan Pendidikan Matematika dalam Menyongsong Era Disrupsi”. Seminar ini telah dilaksanakan di FMIPA Universitas Negeri Malang (UM) pada Selasa, 27 Oktober 2020 secara virtual. Peserta seminar terdiri dari mahasiswa, dosen, dan guru dari berbagai daerah di Indonesia.

Artikel-artikel yang dimuat dalam prosiding ini telah melalui proses seleksi, reviu oleh para ahli bidang matematika dan pendidikan matematika, revisi oleh penulis dan reviu final oleh reviewer untuk menjamin kualitas artikel yang dimuat dalam prosiding ini. Artikel yang dimuat meliputi Topik Pendidikan Matematika seperti: (1) Proses Berpikir, (2) Penerapan Model Pembelajaran, (3) Penelitian Tindakan Kelas (PTK), dan (4) Pengembangan Media. Sedangkan topik Matematika dan Penerapannya mencakup (1) Aljabar, (2) Statistika, (3) Analisis, (4) Terapan, (5) Geometri dan Graph, dan (6) Komputasi. Oleh karena itu, prosiding ini dapat dijadikan sebagai rujukan pengetahuan yang berkualitas.

Kami mengucapkan terimakasih kepada semua panitia dan reviewer kegiatan Seminar Nasional Matematika dan Pembelajarannya yang tidak dapat kami sebutkan semua. Ucapan terimakasih juga kami sampaikan kepada:

1. Prof. Dr. Marsigit, M.A (Universitas Negeri Yogyakarta), selaku pembicara utama.
2. Prof. Dr. Toto Nusantara, M.Si (Universitas Negeri Malang), selaku pembicara utama.
3. Dr. Susiswo, M.Si, ketua Jurusan Matematika FMIPA UM.
4. Pihak-pihak lain yang tidak dapat disebutkan satu per-satu.

Untuk kesempurnaan prosiding pada seminar-seminar selanjutnya, kritik atau saran yang membangun sangat kami harapkan. Akhirnya, semoga prosiding ini dapat memberikan manfaat bagi peserta seminar dan pembaca secara umum.

Dekan FMIPA
Universitas Negeri Malang

Prof. Dr. Hadi Suwono, M.Si



DAFTAR ISI

KATA PENGANTAR..... ii

DAFTAR ISI iii

ARTIKEL PEMBICARA UTAMA..... 1

**PERAN MATEMATIKA DAN PENDIDIKAN MATEMATIKA DALAM MENYONGSONG ERA
DISRUPSI..... 2**

ANTI FUZZY GRAPH..... 14

ARTIKEL PEMBICARA UNDANGAN..... 26

MEMBACA PIKIRAN INTUITIF DAN ANALITIS DARI SOLUSI MATEMATIKA 27

ARTIKEL PEMBICARA PARALEL 34

**ISOMORFISMA GRUP ANTARA HASILKALI LANGSUNG EKSTERNAL DAN HASIL KALI PUSAT
INTERNAL..... 35**

PENGEMBANGAN KRITERIA MODUL INJEKTIF MENJADI MODUL Σ –INJEKTIF..... 41

**KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL MAIN KI DAN
IMPLEMENTASINYA DALAM MATERI PENGUMPULAN DAN PENGOLAHAN DATA DI
SEKOLAH DASAR..... 50**

**KAJIAN ETNOMATEMATIKA DALAM PERMAINAN TRADISIONAL SAMBUNG KAKI DAN
IMPLEMENTASINYA DALAM PEMBELAJARAN TOPIK TEOREMA PHYTAGORAS DAN
KESEBANGUNAN 59**

**EKSPLORASI ETNOMATEMATIKA PADA PERMAINAN PENTENG KHAS MADURA DAN
IMPLIKASINYA DALAM PEMBELAJARAN TOPIK KELIPATAN BILANGAN 69**

**EKSPLORASI ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL BABANGA DARI
KALIMANTAN SELATAN DAN PENGGUNAANNYA DALAM PEMBELAJARAN 79**

**KAJIAN ETNOMATEMATIKA PADA PERMAINAN BASASAKOLAHAN DAN
IMPLEMENTASINYA UNTUK PEMBELAJARAN MATERI PELUANG 88**

**KAJIAN ETNOMATEMATIKA PADA PERMAINAN TAK-TIK BALI DAN IMPLEMENTASINYA
DALAM PEMBELAJARAN PEMODELAN GERAK PARABOLA 97**

**EKSPLORASI ETNOMATEMATIKA PERMAINAN TRADISIONAL “MAIN CEPU” DARI LAMPUNG
DAN IMPLEMENTASI DALAM PEMBELAJARAN MATEMATIKA MATERI FILLING SLOT DAN
COMBINASI 107**

**KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL POKIBO DAN
PENERAPANNYA DALAM RANCANGAN PEMBELAJARAN PELUANG..... 115**

**KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL BALUBUK DARI
KALIMANTAN SELATAN DAN IMPLEMENTASINYA TERHADAP PEMBELAJARAN
MATEMATIKA DENGAN MATERI OPERASI BILANGAN BULAT..... 124**

KAJIAN ETNOMATEMATIKA PADA PERMAINAN CONGKAK DARI BENGKULU DAN IMPLEMENTASINYA PADA PENGEMBANGAN SOAL HOTS.....	135
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TODO-TODO MINYA DAN PENERAPANNYA DALAM PEMBELAJARAN MATEMATIKA UNTUK MATERI PELUANG DI SMP.....	144
KAJIAN ETNOMATEMATIKA PADA PERMAINAN KEPALA BABI DAN IMPLEMENTASINYA DALAM PEMBELAJARAN MATEMATIKA PADA KONSEP PERKALIAN.....	153
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL BAKARAT DAN IMPLEMENTASINYA PADA PEMBELAJARAN TOPIK PERBANDINGAN	160
KAJIAN ETNOMATEMATIKA PADA PERMAINAN BUTA LELE DAN INTEGRASINYA DALAM RANCANGAN PEMBELAJARAN MATEMATIKA TOPIK PERKALIAN KELAS 2 SEKOLAH DASAR	171
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL “MA’ BEKEL” DARI KALIMANTAN TIMUR DAN PENERAPANNYA DALAM PEMBELAJARAN PADA KONSEP PERKALIAN DAN PEMBAGIAN KELAS VI SD.....	181
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL TENGGOH-TENGGOHAN DARI SUMATERA SELATAN DAN PENERAPANNYA PADA MATERI PELUANG DALAM PEMBELAJARAN MATEMATIKA.....	190
KAJIAN ETNOMATEMATIKA PADA PERMAINAN BABULANAN DAN IMPLEMENTASINYA DALAM PEMBELAJARAN MATEMATIKA MATERI GEOMETRI DIMENSI DUA.....	198
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL BABURUNGAN DAN IMPLEMENTASI PADA PEMBELAJARAN MATEMATIKA TOPIK PELUANG.....	208
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL MADEPA-DEPAAN DAN IMPLEMENTASINYA TERHADAP RENCANA PEMBELAJARAN MATERI PENGUKURAN PANJANG	218
EKSPLORASI ETNOMATEMATIKA DALAM PERMAINAN TRADISIONAL CIK-KECIKAN DARI MADURA DAN IMPLEMENTASINYA PADA PEMBELAJARAN MATERI PELUANG.....	225
KAJIAN ETNOMATEMATIKA PADA PERMAINAN KAPAL-KAPALAN DARI BALI DAN IMPLEMENTASINYA UNTUK PEMBELAJARAN MATERI PELUANG	234
KAJIAN ETNOMATEMATIKA DALAM PERMAINAN CEPLI DAN IMPLEMENTASINYA DALAM PEMBELAJARAN PELUANG.....	241
KAJIAN ETNOMATEMATIKA PERMAINAN KUCING-KUCING DAN IMPLEMENTASINYA DALAM MENGENALKAN OPERASI MODULO.....	249
EKSPLORASI ETNOMATEMATIKA PADA PERMAINAN MACAN-MACANAN DARI BALI DAN IMPLEMENTASINYA PADA SOAL HOTS MATEMATIKA TINGKAT SMP DAN SMA	257
KAJIAN ETNOMATEMATIKA PERMAINAN TRADISIONAL MAKUCUNG-KUCUNGAN BALI DAN IMPLEMENTASINYA PADA PEMBELAJARAN MATERI MODULO	268
KAJIAN ETNOMATEMATIKA PADA PERMAINAN TRADISIONAL MAGEMBLUNG DARI BALI DAN IMPLEMENTASINYA DALAM PEMBELAJARAN MATEMATIKA TOPIK OPERASI HITUNG	277



DIMENSI K-METRIK PADA GRAF COCKTAIL PARTY, GRAF BANANA TREE, DAN GRAF LINTASAN KORONA LINTASAN	287
PENERAPAN CAPACITATED VEHICLE ROUTING PROBLEM WITH TIME WINDOWS UNTUK MENYELESAIKAN MASALAH PENDISTRIBUSIAN MANGGA PODANG DI KABUPATEN KEDIRI	295
DIMENSI K-METRIK PADA GRAF MUSICAL, GRAF TURAN, DAN GRAF CYCLE CORONA GRAF COMPLETE BIPARTITE	303
PELABELAN TOTAL SISI TRIMAGIC SUPER PADA GRAF BUTTERFLY	315
PELABELAN HARMONIS PADA GRAF TANGGA SEGI EMPAT VARIASI	320
DIMENSI K-METRIK PADA GRAF FIRECRACKER DAN GRAF BROOM	325
ANALISA PERBANDINGAN INFERENSI FUZZY TSUKAMOTO DAN MAMDANI DALAM MEMPREDIKSI JUMLAH PENDAFTAR MAHASISWA BARU PRODI TADRIS MATEMATIKA IAIN KEDIRI	331
KEKUATAN SISI REFLEKSIF PADA GRAF TANGGA.....	339
RULE BASED PADA FUZZY INFERENCE SYSTEM METODE TSUKAMOTO DALAM SISTEM PENDUKUNG KEPUTUSAN	343
IDENTIFIKASI PENALARAN KREATIF DAN IMITATIF SISWA DITINJAU DARI PERBEDAAN GENDER	362
IMPLEMENTASI TEORI VYGOTSKY DALAM PROJECT BASED LEARNING PADA MATA KULIAH KAJIAN & PENGEMBANGAN BAHAN AJAR.....	370
ANALISIS PENGETAHUAN PROSEDURAL MAHASISWA PADA POKOK BAHASAN UKURAN LETAK DATA BERBANTUAN MINITAB DAN SPSS	379
PENERAPAN VIDEO PEMBELAJARAN BERBANTUAN MINITAB PADA PEMBELAJARAN DARING UNTUK MENINGKATKAN PEMAHAMAN KONSEP ANALISIS REGRESI NON LINIER MAHASISWA	389
PERSEPSI MAHASISWA MATEMATIKA TERHADAP PERKULIAHAN EKONOMETRIKA BERBASIS KELAS VIRTUAL BERBANTUAN MINITAB	399
PROSES BERPIKIR KONSEPTUAL SISWA SMP DALAM MENYELESAIKAN PERMASALAHAN ALJABAR.....	408
PENGARUH DUKUNGAN ORANG TUA DAN KEPERCAYAAN DIRI SISWA TERHADAP HASIL BELAJAR MATEMATIKA MELALUI MEDIASI MOTIVASI BELAJAR SISWA	414
PEMAHAMAN KONSEP REGRESI LINEAR BERGANDA BERBANTUAN MINITAB PADA PERKULIAHAN DARING EKONOMETRIKA MAHASISWA JURUSAN MATEMATIKA	422
TRANSLASI REPRESENTASI VERBAL KE GRAFIK DALAM MENYELESAIKAN SOAL BANGUN DATAR DITINJAU DARI GAYA KOGNITIF PESERTA DIDIK.....	432
PENERAPAN MEDIA SUDUT TRIGONOMETRI (SUTRI) UNTUK MENINGKATKAN PEMAHAMAN DAN KEAKTIFAN SISWA PADA MATERI TRIGONOMETRI.....	441
ANALISIS PROSES BERPIKIR ALJABAR MAHASISWA CALON GURU MATEMATIKA DENGAN KEMAMPUAN MATEMATIKA RENDAH MENURUT TAKSONOMI SOLO.....	448



ANALISIS KESALAHAN REPRESENTASI SIMBOLIK CALON GURU MATEMATIKA DALAM MENYELESAIKAN SOAL KALKULUS	455
KONEKSI MATEMATIS SISWA DALAM MENYELESAIKAN SOAL NON RUTIN SISTEM PERSAMAAN LINEAR TIGA VARIABEL (SPLTV).....	465
ANALISIS KEMAMPUAN BERNALAR MATEMATIS SISWA PADA MATERI TRIGONOMETRI	475
PENERAPAN PEMBELAJARAN BERBASIS INKUIRI BERBANTUAN GAMES SOAL UNTUK MENINGKATKAN MINAT BELAJAR MATEMATIKA SISWA KELAS X	485
PERSEPSI MAHASISWA JURUSAN MATEMATIKA TERHADAP PERKULIAHAN DARING DI MATA KULIAH EKONOMETRIKA PADA POKOK BAHASAN AUTOKORELASI.....	493
PEMODELAN PENDAPATAN ASLI DAERAH ISTIMEWA YOGYAKARTA MENGGUNAKAN REGRESI DATA PANEL	500
MODEL GEOSPASIAL : UNIVERSAL KRIGING DENGAN SEMIVARIOGRAM COPULA	509
PEMODELAN DISCRETE TIME MARKOV CHAIN SUSCEPTIBLE INFECTED RECOVERED (DTMC SIR) PADA PENYEBARAN PENYAKIT EBOLA DI SIERRA LEONE.....	516
PEMODELAN TINGKAT KEMISKINAN DI INDONESIA.....	525
MODEL PEMBELAJARAN PBL DENGAN BERBANTUAN SPSS UNTUK MEMUDAHKAN PEMAHAMAN MAHASISWA PADA MATERI SKEWNESS DAN KURTOSIS.....	532
PENERAPAN REGRESI HURDLE NEGATIVE BINOMIAL (HNB) UNTUK MEMODELKAN DATA JUMLAH KEMATIAN IBU HAMIL DI PROVINSI JAWA TIMUR.....	541
PENGGUNAAN K-MEANS CLUSTERING UNTUK MENGATASI IMBALANCE DATA DENGAN ELM (EXTREME LEARNING MACHINE) SEBAGAI CLASSIFIER.....	548
PEMODELAN AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) PADA JUMLAH KUNJUNGAN WISATAWAN DOMESTIK DI KOTA MALANG TAHUN 2014 - 2018.....	555
PREDIKSI VOLUME PRODUKSI AIR BERSIH PERUSAHAAN DAERAH AIR MINUM (PDAM) KABUPATEN MALANG MENGGUNAKAN METODE EXPONENTIAL SMOOTHING	566
ESTIMASI PARAMETER MODEL REGRESI NONPARAMETRIK SPLINE TRUNCATED DENGAN PENALIZED LEAST SQUARE.....	575

PENGGUNAAN K-MEANS CLUSTERING UNTUK MENGATASI *IMBALANCE DATA* DENGAN ELM (*EXTREME LEARNING MACHINE*) SEBAGAI *CLASSIFIER*

Umi Mahdiyah^{1,a)}, Nalsa Cintya Resti²⁾, Patmi Kasih¹⁾

¹⁾Universitas Nusantara PGRI Kediri

²⁾Institut Agama Islam Negeri Kediri

^{a)}umimahdiyah@gmail.com

Abstrak

Masalah *imbalanced data* (data yang tidak seimbang) selama ini masih menjadi masalah yang cukup penting pada masalah klasifikasi. Data real pada kehidupan sehari-hari umumnya bersifat *imbalanced*, sehingga untuk melakukan klasifikasi butuh proses yang optimal untuk mendapatkan hasil yang optimal pula. Karena data yang tidak seimbang menyebabkan proses klasifikasi menjadi tidak optimal. Pada artikel ini dilakukan proses *undersampling* dengan memanfaatkan algoritma clustering yaitu K-Means. Selanjutnya, data yang sudah dilakukan proses *undersampling* dimasukkan pada proses klasifikasi menggunakan *Extreme Learning Machine*. Data yang digunakan merupakan *benchmark data set* yang bersifat *imbalanced*. Hasil yang didapatkan dalam penelitian ini cukup baik, dilihat dari nilai rata-rata *precision*, *recall*, dan akurasi.

Kata kunci: *Imbalanced Data*, *Undersampling*, *K-Means*, ELM

Pendahuluan

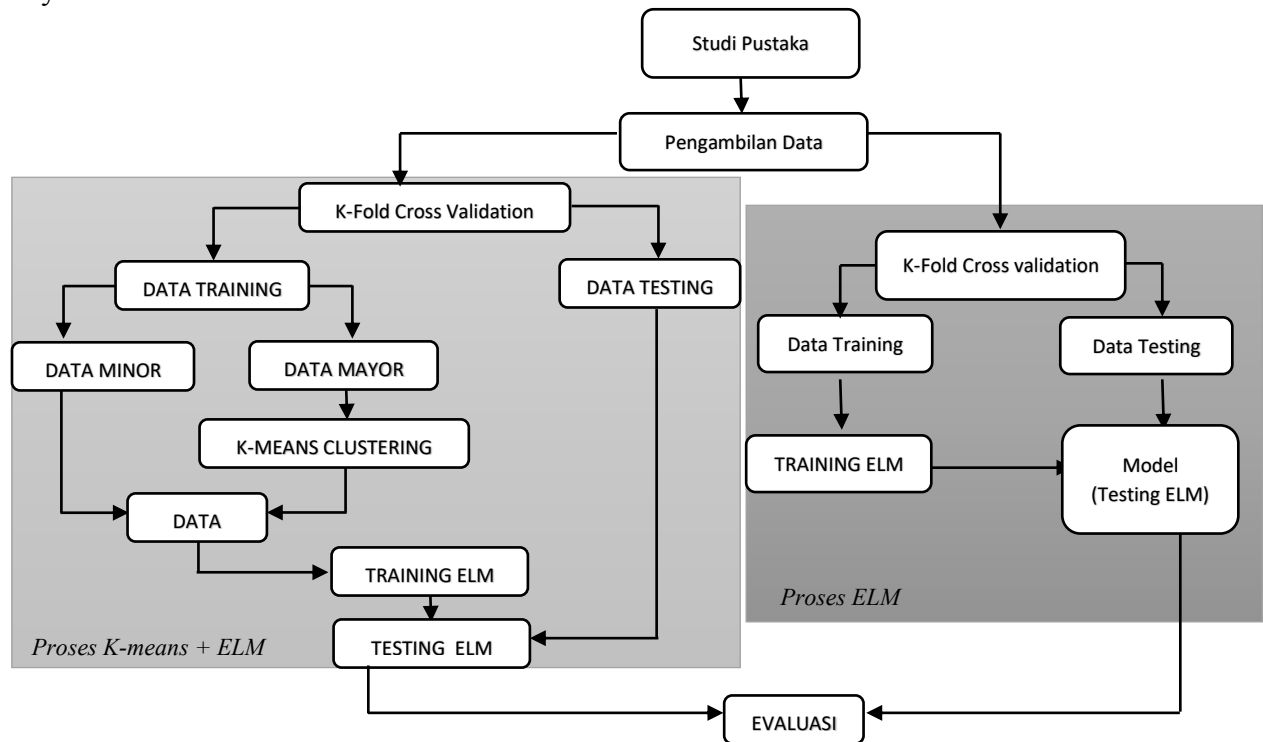
Imbalanced data merupakan salah satu permasalahan yang cukup krusial dalam proses klasifikasi. Karena dengan data yang berkarakter *imbalanced* atau tidak seimbang maka hal tersebut dapat mempengaruhi performansi dari suatu algoritma klasifikasi. Akurasi akan sangat baik untuk kelas mayoritas, akan tetapi akan buruk untuk data minoritas. Permasalahan ketidakseimbangan kelas dapat ditangani dengan 2 pendekatan, yaitu pendekatan level data dan pendekatan level algoritma [1].

Saat ini penanganan masalah *imbalanced data* sudah cukup banyak dilakukan, diantaranya menggunakan *undersampling* maupun *oversampling* serta pengembangannya [2]. Dalam penelitian ini *undersampling* dipilih untuk menyelesaikan masalah *imbalanced data* dengan tujuan supaya lebih ringan proses komputasinya, selain itu supaya waktu yang dibutuhkan bisa lebih cepat. K-Means merupakan salah satu algoritma clustering yang sering digunakan karena performansinya yang bagus. Dalam penelitian ini *K-means clustering* digunakan dalam proses *sampling* untuk mengatasi masalah *imbalanced data*. *Extreme Learning Machine* dipilih untuk classifier, sehingga untuk mempermudah penyebutan proses *K-means clustering* yang digabung dengan proses ELM dapat disederhakan KM-ELM.

Metode

Langkah-langkah dan proses penelitian digambarkan pada Gambar 1. Langkah awal melakukan studi literatur terkait data yang tidak seimbang. Selanjutnya dilakukan pengambilan data pada web <http://www.keel.es/>. Data yang digunakan dalam penelitian ini merupakan data sekunder yang dapat digunakan banyak peneliti untuk melakukan eksperimen karena data pada web tersebut bersifat *open data*. Detail data pada penelitian ini dapat dilihat pada Tabel 1, data yang digunakan pada penelitian ini merupakan data biner. Data tersebut kemudian dibagi menjadi 2 bagian yaitu data training dan data testing, dengan menggunakan *5 fold cross validation*.

Data training adalah data yang digunakan untuk melatih sistem. Dalam penelitian ini data yang digunakan adalah data biner, sehingga data dibagi lagi menjadi data minor dan data mayor. Selanjutnya pada data mayor dilakukan proses sampling dengan menggunakan K-means clustering [3]. Tujuan dari sampling dengan untuk meringkas data berdasarkan kesamaan karakter sehingga data minor dan data mayor tidak memiliki selisih yang sangat banyak.



Gambar 1. Diagram Alir Penelitian

Gambar 1 merupakan diagram alir dari penelitian yang dilakukan. Pada Gambar 1 dapat dilihat, bahwa setelah proses pengambilan data, data tersebut diberikan 2 perlakuan berbeda, pada bagan sebelah kanan dilakukan proses ELM biasa. Pada bagan sebelah kiri data training dibedakan menjadi 2 yaitu data mayor dan data minor. Pada data mayor dilakukan proses *undersampling* menggunakan *K-means clustering* dengan *k* dipilih sesuai perbandingan data awal (data minor dan mayor) sehingga diharapkan dapat mengatasi masalah *imbalanced data* untuk data training, sehingga data minor dan data mayor dapat ditemukan polanya dengan lebih baik pada sistem. Selanjutnya dilakukan proses training dan testing.

Setelah proses training dan testing menggunakan ELM biasa serta ELM dengan proses *undersampling* menggunakan *k-means clustering* selesai, selanjutnya dilakukan evaluasi berupa perbandingan hasil dari 2 perlakuan yang telah dilakukan. Evaluasi pada penelitian ini dilakukan dengan menggunakan perhitungan pada *confusion matrix*.

Tabel 1 merupakan tabel data yang digunakan dalam penelitian, karena yang digunakan **data biner** maka dibuat dua nama yaitu data positif dan data negatif. Data yang digunakan tidak semuanya dalam bentuk data *imbalanced* yang ekstrim, artinya ada yang perbandingan data negatif dan positif adalah 1:2 tetapi datanya cukup besar sehingga selisih jumlah data negatif dan positif cukup banyak. Data yang tidak *imbalanced* juga ikut diujikan dengan tujuan mengecek apakah metode yang dilakukan berlaku untuk semua jenis data.

Tabel 1. Data Yang Digunakan Pada Penelitian

No	Nama Data	Banyak Data		Perbandingan Data
		Data Negatif	Data Positif	
1	Pima	150	134	1:1
2	Ecoli	129	39	3:1
3	Abalone	344	21	16:1
4	Glass1	69	38	2:1
5	Haberman	113	41	3:1
6	Yeast	527	215	2:1
7	Spambase	1394	906	2:1
8	Cleveland	80	6	13:1
9	Dermatology	169	10	17:1

Menurut Daniel dan Eko [4], Langkah-langkah algoritma K-Means adalah sebagai berikut:

- Pilih secara acak k buah data sebagai pusat cluster.
- Jarak antara data dan pusat cluster dihitung menggunakan Euclidian Distance. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D_{i,j} = \sqrt{(x_{1i}-x_{1j})^2 + (x_{2i}-x_{2j})^2 + \dots + (x_{ki}-x_{kj})^2}$$

dimana:

D (i, j) = Jarak data ke i ke pusat cluster j

x_{ki} = Data ke i pada atribut data ke k

x_{kj} = Titik pusat ke j pada atribut ke k

- Data ditempatkan dalam cluster yang terdekat, dihitung dari tengah cluster.
- Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam cluster

Misalkan untuk data glass terdapat 9 atribut dengan data awal ada 107 data, untuk deskripsi data tersebut bisa dilihat pada <https://www.kaggle.com/uciml/glass?select=glass.csv>, selanjutnya data dibagi menjadi 2 data minor dan mayor. Data minor sebanyak 38 dan data mayor sebanyak 69, selanjutnya dari data tersebut dipilih data mayor untuk dilakukan undersampling. Misalkan dipilih 2 *centroid* dari data mayor karena akan membagi data menjadi 2 cluster.

Tabel 2. Contoh Data Glass

Data ke	RI	Na	Mg	Al	Si	K	Ca	Ba	Fr
1	1.5184	14.321	3.2597	2.2192	71.249	1.4593	5.7851	1.6285	0
2	1.5193	13.197	3.3316	1.2787	72.358	0.6023	9.1422	0	0.0561
3	1.519	13.583	3.3495	1.2305	72.07	0.58995	8.9055	0	0
dst									

Misalkan dipilih 2 centroid dari data tersebut:

Centroid 1: 1.5165 13.038 3.3989 1.2594 73.008 0.52164 8.5827 0 0
 Centroid 2: 1.5176 12.712 3.4214 1.1984 73.198 0.58995 8.6365 0 0

Selanjutnya hitung jarak antara data dan pusat cluster dihitung menggunakan *Euclidian Distance*.

Misal untuk data 1: jarak data 2 dengan *centroid 1*, sehingga $i=1, j=2, k=9$ karena banyak atributnya adalah 9.

$$D_{2,1} = \sqrt{(x_{12}-x_{11})^2 + (x_{22}-x_{21})^2 + \dots + (x_{ki}-x_{kj})^2}$$

$$D_{2,1} = \sqrt{(1.5193-1.5165)^2 + (14.321-13.038)^2 + \dots + (0.0561-0)^2}$$

$$= 5,679122$$

Selanjutnya dengan cara yang sama juga hitung jarak data ke *centroid 2*.

Sehingga dapat dibuat tabel 3:

Tabel 3. Contoh Perhitungan Jarak ke centroid

Data	Jarak		Lebih dekat ke centroid-
	Centroid 1	Centroid 2	
1	76,69098	77,23831	1
2	80,6786	81,30027	1
dst			

Jika sudah dibuat perhitungan jarak ke *centroid* maka selanjutnya adalah update centroid, dengan cara hitung seluruh rata2 fitur data yang masuk d cluster 1 maka akan menjadi *centroid* pengganti dari centroid yang sebelumnya. Selanjutnya, dihitung kembali jarak setiap data dengan *centroid* yang sudah diupdate. Iterasi dihentikan jika anggota cluster sudah tidak berubah.

Setelah data sudah di cluster dengan menggunakan *K-means* selanjutnya diambil beberapa bagian data pada setiap cluster disesuaikan dengan perbandingan data tak seimbang. Setelah itu data yang sudah diambil dari data mayor digabung lagi dengan data minor, kemudian data gabungan dimasukkan ke dalam algoritma *Extreme Learning Machine*.

Konsep utama dari ELM seperti yang disajikan dalam *paper* Huang adalah sebagai berikut:

Diberikan *training set* $\mathfrak{N} = \{(x_j, t_j) | x_j \in R^{n \times m}, t_j \in R^n, j \in [1, N]\}$, fungsi aktivasi $g(x)$, dan bilangan *hidden node* \tilde{N}

Step 1: masukkan secara random bobot w_i dan bias $b_i, i \in [1, \tilde{N}]$

Step 2: hitung *output* matriks *hidden layer* H

Step 3: hitung bobot *output* β

$$\beta = H^\dagger T$$

dengan $T = [t_1, \dots, t_N]^T$,

t_1, \dots, t_N didapat dari data training $\mathfrak{N} = \{(x_j, t_j) | x_j \in R^{n \times m}, t_j \in R^n, j \in [1, N]\}$

H^\dagger adalah *Generalized Inverse*

$$H^\dagger = (H^T H)^{-1} H^T \text{ atau,}$$

$$H^\dagger = H^T (H H^T)^{-1}$$

Setelah dilakukan proses training dan testing, selanjutnya dilakukan evaluasi dengan menggunakan *confusion matriks*. Untuk mengukur performa dari klasifikasi data pengujian

imbalance dalam hal ini digunakan *confusion matrix*, *precision* dan *recall*, *specificity*, dan *G-mean* [5]. Tabel *confusion matrix* Dapat dilihat pada tabel 4. Dalam pembuatan dan pengujian sistem melibatkan software, software yang digunakan adalah Matlab R2015a, sedangkan processor PC yang digunakan adalah intel core i5, dengan RAM 4GB.

 Tabel 4. *Confusion Matrix*

		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Prediksi	<i>True</i>	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
	<i>False</i>	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Precision adalah presentase dari data yang diprediksi benar oleh *classifier* yang bernilai benar.

$$precision = \frac{TP}{TP + FP}$$

Recall adalah porsi dari data sampel yang diprediksi benar oleh *classifier*.

$$recall = \frac{TP}{TP + FN}$$

$$sensitifity = recall$$

Geometric mean telah digunakan beberapa peneliti untuk mengevaluasi *classifier* pada dataset yang *imbalanced*. *G-mean* mengindikasikan keseimbangan antara kinerja klasifikasi pada kelas mayoritas dan minoritas. Ukuran *G-mean* diambil berdasarkan *sensitifity* (akurasi dari data positif) dan *specificity* (akurasi data negatif).

$$specificity = 1 - \frac{FP}{FP + TN}$$

$$G - mean = \sqrt{sensitifity \times specificity}$$

Hasil dan Pembahasan

Hasil dari penelitian dapat dilihat pada Tabel 5. Dari tabel tersebut dapat dilihat bahwa setelah dilakukan sampling pada data yang berkarakter *imbalanced* pada saat proses training dapat meningkatkan performansi *Extreme Learning Machine* untuk kasus data yang perbandingan data positif dan negatifnya cukup besar karena dalam penelitian ini diujikan untuk klasifikasi biner. Untuk kasus data yang perbandingan data negatif dan positifnya tidak terlalu jauh nilai *precision*, *recall* dan *specificity* meningkat hanya saja tidak meningkat tinggi, bahkan untuk *specificity* pada data Pima, glass1, dan Yeast terjadi penurunan sedikit, artinya untuk data Pima, Glass1, dan Yeast ELM dapat lebih baik mengenali data negatif daripada KM-ELM. Sedangkan untuk data lain KM-ELM nilai *specificity*nya di atas ELM biasa artinya untuk data yang perbandingan data masing-masing kelasnya lebih dari 1:3 KM-ELM dapat lebih baik *specificity* nya dibanding ELM.

Nilai *precision*, *recall* dan *G-mean* untuk semua data, baik data yang perbandingan data masing-masing kelasnya tinggi maupun rendah tetap nilai *precision*, *recall* dan *G-mean* lebih tinggi dibandingkan ELM biasa. Dapat dilihat juga pada tabel 5 bahwa untuk data yang perbandingan data kelasnya rendah nilai *precision*, *recall* dan *G-mean* hanya naik sedikit saja, untuk data yang perbandingannya tinggi kenaikannya cukup drastis, misal untuk data Cleveland, pada data ini ELM tidak mengenali sama sekali data positifnya, setelah dilakukan sampling dengan *K-Means* pada proses training data positif dapat dikenali dengan sempurna.

Untuk nilai akurasi data dengan ELM lebih tinggi dibanding KM-ELM karena pada akurasi kebanyakan yang dikenali ada pada data mayoritas, nilai akurasi juga akan tetap tinggi meski kelas minoritas pada data tersebut tidak dikenali sama sekali, serta data mayoritas bisa dikenali semua.

Tabel 5. Tabel Hasil Penelitian

Nama Data	Akurasi		Precision		Recall		Specificity		G-mean		CPU Time	
	ELM	KM-ELM	ELM	KM-ELM	ELM	KM-ELM	ELM	KM-ELM	ELM	KM-ELM	ELM	KM-ELM
Pima	0,63	0,65	0,48	0,68	0,52	0,63	0,69	0,68	0,60	0,65	0,0101	1,18
Ecoli	0,84	0,92	0,62	0,97	0,82	0,85	0,85	0,98	0,83	0,91	0,0114	1,07
Abalone	0,94	0,86	0,50	1,00	0,10	0,71	0,99	1,00	0,31	0,85	0,0125	1,276
Glass1	0,64	0,54	0,50	0,65	0,21	0,29	0,88	0,82	0,43	0,49	0,0094	0,2028
Haberman	0,74	0,67	0,53	0,82	0,24	0,45	0,92	0,89	0,47	0,63	0,01	0,197
Yeast	0,73	0,71	0,59	0,81	0,20	0,47	0,94	0,91	0,44	0,66	0,0218	1,95
Spambase	0,69	0,73	0,59	0,88	0,66	0,61	0,71	0,89	0,68	0,74	0,2714	4,829
Cleveland	0,93	1,00	NaN	1,00	0	1,00	1,00	1,00	0	1,00	0,00074	0,215
Dermatology	0,96	0,95	0,71	1,00	0,50	0,90	0,99	1,00	0,70	0,95	0,0115	1,198
Rata-rata	0,79	0,78	0,56	0,87	0,36	0,66	0,89	0,91	0,50	0,76	0,03987	1,3464

*ELM = Extreme Learning Machine

*KM-ELM = sebelum proses ELM data terlebih dahulu dilakukan sampling dengan K-means

Kesimpulan dan Saran

Dari uraian pada hasil dan pembahasan dapat disimpulkan bahwa melakukan proses sampling dengan K-Means sebelum dilakukan proses klasifikasi dengan ELM dapat meningkatkan performansi dari ELM itu sendiri pada kasus data yang bersifat *Imbalanced*. Untuk kasus data yang bukan bersifat *imbalanced* sebaiknya bisa menggunakan algoritma lain untuk klasifikasi supaya mendapatkan hasil yang lebih optimal.

Daftar Rujukan

- [1] B. Santoso, H. Wijayanto, K. A. Notodiputro dan B. Sartono, “Class imbalanced problems: a review,” dalam *Conference Series: Earth and Environmental Science*, -, 2017.
- [2] U. Mahdiyah, M. I. Irawan dan E. M. Imah, “Integrating data selection and extreme learning machine for imbalanced data,” *Procedia Computer Science* , pp. 221-229, 2015.
- [3] P. Arora, D. Deepali dan S. Varshney, “Analysis of K-Means and K-Medoids Algorithm For Big Data,” dalam *International Conference on Information Security & Privacy (ICISP2015)*, India, 2016.
- [4] D. R. Kaparang dan E. Sedyono, “Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis,” *d’Cartesian: Jurnal Matematika dan Aplikasi* , vol. 2, no. 2, pp. 18-25, 2013.
- [5] H. He dan E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Transactions on Knowledge and Data Engineering* , vol. 21, no. 9, pp. 1263-1284, 2009.
- [6] P. Giganti, “Why Teach Problem Solving, Part I: The World Needs Good Problem Solvers!.,” *ComMuniCator*, vol. 31, no. 4, pp. 15-16, 2007.
- [7] S. Dewiyani, “Improving Students Soft Skills using Thinking Process Profile Based on Personality Types,” *International Journal of Evaluation and Research in Education (IJERE)*, pp. pp 118-129, 2015.