

Klasifikasi Calon Mahasiswa Bidikmisi dengan Algoritma K-Nearest Neighbor

1st Jajam Haerul Jaman
Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang
jajam.haeruljaman@staff.unsika.ac.id

2nd Syah Adi Fahlevi
Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang
syah.adi1412@student.unsika.ac.id

Abstract—The Bidikmisi scholarship program is a tuition assistance program provided by the Ministry of Technology Research and Higher Education to students who have a lower middle economic level but have good academic potential. In the selection process has not used a special method in accordance with existing rules, less objective, less effective and requires a relatively longer time. K-Nearest Neighbor is a method that uses a supervised algorithm in which the new testing data is classified based on the majority of classes. Optimization of the classification final using KNN to get or know Determine the best K with experiments $k = 1$ to $k = 10$. by producing the best K that is $k = 1$ by producing the highest accuracy value of 97.2067%, the highest precision value of 0.972 and the highest recall value of 0.972 the best target k is $k = 1$.

Abstrak—Program beasiswa bidikmisi merupakan program bantuan biaya pendidikan yang diberikan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi kepada mahasiswa yang memiliki tingkat ekonomi menengah ke bawah namun memiliki potensi akademik yang baik. Dalam proses penyeleksiannya belum menggunakan sebuah metode khusus dalam pemilihan sesuai dengan aturan yang ada. kurang objektif, kurang efektif dan membutuhkan waktu yang relatif lebih lama. *K-Nearest Neighbor* adalah suatu metode yang menggunakan algoritma *supervised* dimana *data testing* yang baru diklasifikasikan berdasarkan mayoritas kelas.. Optimasi klasifikasi mahasiswa menggunakan *K-Nearest Neighbor* untuk mendapatkan atau mengetahui Menentukan K terbaik dengan percobaan $k = 1$ sampai $k = 10$. dengan menghasilkan K terbaik yaitu $k = 1$ dengan menghasilkan nilai akurasi tertinggi sebesar 97.2067 %, nilai *precision* tertinggi sebesar 0.972 dan nilai *recall* tertinggi sebesar 0.972 target k terbaik yaitu $k = 1$.

Kata Kunci : Bidikmisi, Klasifikasi, *K-nearest-Neighbor*

I. PENDAHULUAN

Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Bab V pasal 12 (1.c), menyebutkan bahwa setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan beasiswa bagi yang berprestasi yang orang tuanya tidak mampu membiayai pendidikannya. Program beasiswa pendidikan

mahasiswa dengan keterbatasan ekonomi namun berprestasi atau yang sering dikenal dengan sebutan bidikmisi merupakan program bantuan biaya pendidikan yang diberikan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi kepada calon mahasiswa yang memiliki tingkat ekonomi menengah ke bawah namun memiliki potensi akademik yang baik. Oleh karena itu, tidak sembarang mahasiswa yang bisa mendapatkan beasiswa tersebut. Hanya mahasiswa tertentu saja yang dinyatakan memenuhi persyaratan dan kriteria yang telah ditetapkan langsung oleh Kementerian Riset Teknologi dan Pendidikan Tinggi sebagai mahasiswa yang berhak dan layak menerima bidikmisi. Seperti yang tertera pada petunjuk teknis pelaksanaan program beasiswa bidikmisi semua penerima beasiswa bidikmisi diwajibkan mengikuti aturan yang telah disetujui dan tertera sebagai bahan acuan bagi mahasiswa penerima beasiswa bidikmisi untuk tetap bersikap dan mengikuti aturan yang telah ditetapkan Aturan yang telah ditetapkan itu bersifat mengikat, karena sudah adanya perjanjian dengan kedua belah pihak

II. RUMUSAN MASALAH

Bagaimana mengklasifikasi calon penerima bidikmisi menggunakan algoritma *K-Nearest Neighbor* ?

Bagaimana performa algoritma *K-Nearest Neighbor* dalam mengklasifikasi penerimaan mahasiswa bidikmisi dengan menggunakan *confusion matrix* ?

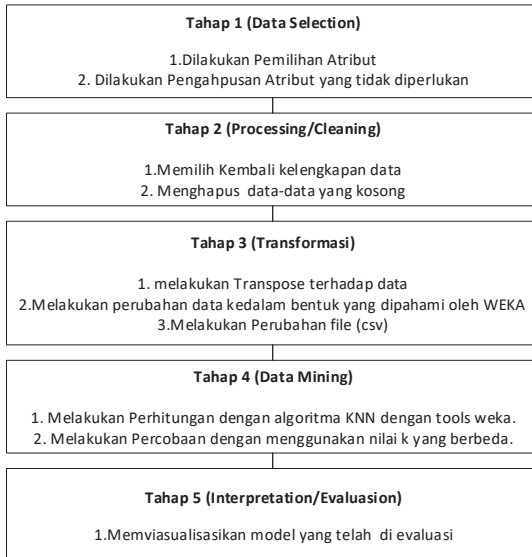
III. TUJUAN PENELITIAN

Mengklasifikasi calon penerima bidikmisi menggunakan algoritma *K-Nearest Neighbor*.

Mengevaluasi performa algoritma *K-Nearest Neighbor* dalam mengklasifikasi penerimaan mahasiswa bidikmisi menggunakan algoritma, dengan menggunakan *confusion matrix*.

IV. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan yaitu *Knowledge discovery in database* (KDD). Dimana tahapannya yaitu: *Selection, Preprocessing/cleaning, Transformasi, Data Mining* Evaluasi.:



Gambar. 1. Alur Penelitian

V. HASIL

A. Data Selection

Pada tahap data *selection* dilakukan pemilihan data yang relevan berdasarkan data yang diperoleh, dikarenakan tidak semua data digunakan. Data yang diperoleh berjumlah 358 data, dalam pemilihan atribut analisis yang digunakan adalah dengan melakukan wawancara kepada pihak terkait yang biasa bertugas sebagai tim penerima bidik misi sehingga didapat beberapa atribut yang penting dan berpengaruh terhadap hasil yaitu: nilai ujian, jumlah saudara, pendapatan orang tua, pekerjaan orang tua dan pendidikan orang tua

B. Data cleaning

Pada tahap ini dilakukan proses pembersihan data calon pelamar penerima bidikmisi tahun 2018. Tahapan yang dilakukan dalam proses *data cleaning* adalah dengan mengecek *missing value* yang ada dalam data calon pelamar penerima bidikmisi. Tahapan selanjutnya adalah mengecek duplikasi data dan mereduksi data sesuai dengan kebutuhan. Dikarenakan data ini lengkap maka jadi tidak dilakukan proses penanganan *missing value*. Selain itu didalam data ini tidak terdapat duplikasi data dan jumlah data yang akan diolah sudah sesuai dengan kebutuhan saat proses *data mining*.

C. Data Transformasi

Pada tahap data *transformation* dilakukan perubah bentuk data menjadi ke dalam bentuk data yang sesuai untuk dilakukan proses *data mining*. Pada tahap ini tidak seluruh kriteria akan digunakan pada proses *data mining*, dikarenakan tidak seluruh kriteria datanya relevan atau tidak memiliki pengaruh pada hasil *data mining*. Tujuan menghilangkan kriteria data yang tidak relevan adalah untuk memudahkan proses klasifikasi. Transformasi dilakukan dengan menggunakan *mean* dan standar deviasi. Proses transformasi yang dilakukan adalah dengan melakukan perubahan tipe data dari data *numeric* menjadi *kategorical*. Adapun kriteria yang di transformasi adalah, nilai ujian, jumlah saudara pendapatan orang tua

1	NILAI	PENDIDIKAN	PEKERJAAN	PENGHASILAN	TANGGUNGAN	CLASS
2	B	Tamat SMTA	Pegawai swasta bukan guru / dosen	A	B	TIDAK DIETRIMA
3	B	Tamat SMTA	Pegawai swasta bukan guru / dosen	B	B	TIDAK DIETRIMA
4	B	Tamat SD	TIDAK DIETRIMA bekerja	A	A	TIDAK DIETRIMA
5	B	Tamat SD	Petani / Nelayan	A	B	DITERIMA
6	B	Tamat SMTA	Wiraswasta / Eksekutif / Pedagang	B	B	TIDAK DIETRIMA
7	B	Tamat SMTA	Buruh	B	B	DITERIMA
8	B	Tamat SMTA	Buruh	A	B	TIDAK DIETRIMA
9	B	TIDAK DIETRIM	Buruh	A	A	DITERIMA
10	B	Tamat SD	Petani / Nelayan	A	B	DITERIMA
11	B	Diploma 3 (D3)	Pensiunan swasta	C	A	TIDAK DIETRIMA
12	B	Tamat SMTA	Pegawai swasta bukan guru / dosen	C	A	TIDAK DIETRIMA
13	B	Tamat SD	Buruh	A	B	TIDAK DIETRIMA
14	B	Tamat SMTA	Buruh	C	B	TIDAK DIETRIMA
15	B	Tamat SMTA	Wiraswasta / Eksekutif / Pedagang	B	B	TIDAK DIETRIMA
16	B	Sarjana (S1)	TIDAK DIETRIMA bekerja	A	A	TIDAK DIETRIMA
17	B	Tamat SD	Wiraswasta / Eksekutif / Pedagang	A	B	DITERIMA
18	B	Tamat SD	Wiraswasta / Eksekutif / Pedagang	C	B	DITERIMA
19	B	TIDAK DIETRIM	Buruh	A	A	TIDAK DIETRIMA
20	B	Tamat SMTA	Pegawai swasta bukan guru / dosen	B	B	DITERIMA
21	C	Diploma 2 (D2)	Lainnya	B	B	TIDAK DIETRIMA

Gambar. 2. Sampel Hasil Transformasi

Tabel 1 Nilai Ujian

Nilai Ujian	
A	<536.758
B	<536.758-565.8038
C	>565.8038

Tabel 2 Jumlah Saudara

Jumlah Saudara	
A	<1
B	1 sampai 3
C	>3

Tabel 3 Penghasilan

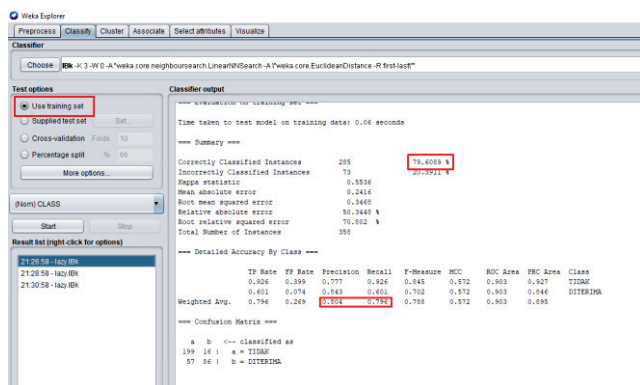
Penghasilan	
A	<1.000.000
B	<1.000.000 – 2.000.000
C	>2.000.000

D. Data Mining

Tahap *data mining* dilakukan pemilihan metode algoritma untuk menemukan pola atau informasi mengklasifikasi data calon penerima bidikmisi Karawang. Pada penelitian ini, penerapan data mining menggunakan algoritma *K-Nearest Neighbor* untuk memproses pengolahan data klasifikasi. Hasil pengolahan data dengan menggunakan algoritma *K-Nearest Neighbor* yang akan diuji dengan *test option* menggunakan *use training set*

Berikut hasil pemodelan algoritma *K-Nearest Neighbor* 2 kategori *class* target menggunakan *tools* WEKA 3.8.3:

1. *K-Nearest Neighbor* dengan $k = 1$



Gambar. 2. Hasil 3 Kategori *K-Nearest Neighbor* dengan $K = 1$

Berdasarkan gambar diatas dapat dilihat bahwa hasil yang diperoleh dari pemodelan *K-Nearest Neighbor* menggunakan *test options - use training set*, persentase untuk *Correctly Classified Instance* adalah sebesar 79.6089 % sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 20.3911 %. Dimana dari 358 data pelamar bidikmisi, ada 285 data berhasil diklasifikasikan dengan benar dan 73 data tidak berhasil diklasifikasikan dengan benar. Pola yang sudah dibuat diuji menggunakan *data testing*, setelah itu menghasilkan *data confusion matrix*.

$$Accuracy = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Total jumlah prediksi yang dilakukan}} \times 100 \%$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN} \times 100\%$$

$$100\% = \frac{199 + 86}{358} = \frac{285}{358}$$

$$= 79.6089\%$$

E. Evaluasi

Pada tahap evaluasi dilakukan antara 2 kategori *class* target yaitu kategori diterima dan tidak diterima, target dari percobaan sebanyak 10 kali dimulai pada $k = 1$ sampai $k = 10$. Hasil percobaan algoritma *K-Nearest Neighbor* dapat dilihat pada Tabel 4.4.

Tabel 4 Hasil Evaluasi

Jumlah K	kategori <i>class</i> target		
	Accuracy	Precision	Recall
K1	79.6089 %	0.804	0.796
K2	70.9497 %	0.715	0.709
K3	69.2737 %	0.694	0.693
K4	66.7598 %	0.675	0.668
K5	65.0838 %	0.657	0.651
K6	64.8045 %	0.656	0.648
K7	64.8045 %	0.658	0.648
K8	63.9665 %	0.654	0.640
K9	63.6872 %	0.653	0.637
K10	63.1285 %	0.647	0.631

Berdasarkan hasil evaluasi pada *data testing*, diketahui hasil nilai tertinggi untuk akurasi, laju *error* dan hasil tersebut akan dibahas lebih lanjut pada tahap selanjutnya. Oleh karena itu tahap evaluasi telah sesuai dengan tahap *data mining* dalam 2 kategori *class* target. Berdasarkan data diatas nilai akurasi tertinggi terdapat pada jumlah $k=1$ dengan nilai akurasi sebesar 79.6089 % sedangkan untuk nilai presisi tertinggi dan *recall* terdapat pada jumlah $k=1$ dengan nilai sebesar 0.804 dan 0.796.

VI. KESIMPULAN

A. Penelitian ini mengklasifikasikan data calon penerima bidikmisi tahun 2018 dengan menggunakan algoritma *K-Nearest Neighbor* dengan jumlah data sebanyak 358 data pelamar bidikmisi. nilai k yang digunakan dalam proses *data mining* yaitu nilai $k=1$ sampai dengan nilai $k=10$. Untuk pembagian *data testing* dan *data training* menggunakan *option test use training set*. Nilai akurasi terbaik hasil dari pengolahan *data mining* dievaluasi menggunakan *confusion matrix* dengan komponen yaitu akurasi, presisi dan *recall*.

B. Evaluasi Algoritma *K-Nearest Neighbor* menggunakan *confusion matrix* diukur berdasarkan nilai *accuracy*, *precision* dan *recall*. Pengujian kinerja menerapkan 5

kriteria yaitu nilai ujian, jumlah saudara, pendapatan orang tua, pekerjaan orang tuadan dan pedidikan orang tua. Menentukan K terbaik dengan percobaan $k = 1$ sampai $k = 10$. K terbaik yaitu $k = 1$ dengan menghasilkan nilai akurasi tertinggi sebesar 79.6089 % dimana hasil nilai akurasi didapat dari jumlah data yang diprediksi secara benar (TP) dibagi total jumlah prediksi yang dilakukan. Nilai *precision* tertinggi sebesar 0.804 dimana nilai tersebut didapat dari jumlah prediksi yang benar (TP) dibagi jumlah dimana jumlah data positif yang terklasifikasi dengan benar (TP) ditambah jumlah data positif namun terklasifikasi salah (FP). Nilai *recall* tertinggi sebesar 0.796 dimana jumlah data positif yang terklasifikasi dengan benar (TP) dibagi jumlah data positif yang terklasifikasi dengan benar (TP) ditambah jumlah data negatif namun terklasifikasi salah (FN) target k terbaik yaitu $k = 1$, dimana dari 358 data pelamar bidikmisi ada 285 data pelamar yang berhasil diklasifikasikan dengan benar dan 73 data pelamar tidak berhasil diklasifikasikan dengan benar.

REFERENCES

- [1] D. B. Kemenristekdikti, PANDUAN PENDAFTARAN BEASISWA BIDIKMISI, Jakarta: Ditjen Belmawa Kemenristekdikti , 2019.
- [2] H. J. Watson and R. H. Sprague, Decision Support Systems: Putting Theory Into Practice, 2013.
- [3] E. Turban, Decision Support Systems and Intelligent Systems, 2012.
- [4] S. R. Siregar and Nurhayati, "Sistem Pendukung Keputusan Pemberian Kredit Pns Dengan Metode AHP dan Topsis (Studi Kasus : Pt. Bank Sumut Cabang Binjai)," 2018.
- [5] Saifulloh and Kusriani, "Sistem Pendukung Pengambilan Keputusan," 2016.
- [6] E. Rahmanita and Y. Kustiyarningsih, "Sistem Pendukung Keputusan Untuk Menentukan Penjurusan SIta Dengan Metode ID3 dan C4.5," 2016.
- [7] R. S. Pressman, Rekayasa Perangkat Lunak, 2012.
- [8] S. Kusumadewi, S. Hartati, A. Harjoko and R. Wardoyo, Fuzzy Multi-Attribute Decision Making (Fuzzy MADM), Yogyakarta: Graha Ilmu, 2006.
- [9] N. Kosasih, Sistem Pendukung Keputusan Pmilihan Motor Dengan Metode Simple Additive Weighting (SAW), 2017.
- [10] A. Kadir, Pengenalan Sistem Informasi Edisi Revisi, 2014.
- [11] E. Ismanto and E. Novera, "Sistem Pendukung Keputusan Penerimaan Karyawan Dengan Metode Simple Additive Weighting (SAW)," 2017.
- [12] Friyadie, "Penerapan Metode Simple Additive Weighting (Saw) Dalam Sistem Pendukung Keputusan Promosi Kenaikan Jabatan," 2016.
- [13] N. I. P. Astuti, Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa, 2018.
- [14] M. H. Alqozhri, Sistem Pendukung Keputusan Pengangkatan Karyawan Tetap menggunakan Algoritma C45, 2015.
- [15] M. I. Adriansyah, B. R. Aditya and W. Wikusna, "APLIKASI BERBASIS WEB PENGOLAHAN NILAI AKADEMIK (STUDI KASUS: SD NEGERI HARAPAN BARU 2 BEKASI)," *e- Proceeding of Applied Science*, 2(3), pp. 824-829, 2016.
- [16] Rosa A.S & M. Shalahuddin, Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek), Bandung: Informatika, 2018.
- [17] B. Prakasa, "Implementasi Big Data Pada Data Transaksi Tiket Elektronik Bus Rapid Transit (BRT)," *CITEE 2017*, pp. 370-376, 2017.
- [18] E. S. Rahayu, R. Satria and C. Supriyanto, "Penerapan metode average gain , threshold pruning dan cost complexity pruning untuk split atribut pada algoritma C4.5," *Jurnal of Intelligent Systems*, pp. 91-97, 2015.
- [19] A. Amrin, "Analisa Kelayakan Pemberian Kredit Mobil Dengan Menggunakan Metode Neural Network Model Radial Basis Function," *Paradigma*, pp. 102-107, 2017.
- [20] A. Diana, "ISSN VOL-12," *Confusion Matrix*, pp. 2-7, 2014.
- [21] H. David J, M. Heikki and S. Padhraic, Principles Of Data Mining, London: MIT Press, 2001.
- [22] W. Ian H, F. Eibe, H. Mark A. and P. Christopher J., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 1999.
- [23] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *Penelitian Ilmu Komputer*, pp. 65-76, 2013.
- [24] R. I. Ndaumanu, Kursini and M. R. Arief, "Analisa Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor," *Jatsi*, pp. 1-15, 2014.
- [25] F. Gorunescu, Data Mining Concepts, Models and Techniques, Heidelberg: Springer, 2011.
- [26] Jaroji, Danuri and F. P. Putra, "K-MEANS UNTUK MENENTUKAN CALON PENERIMA BEASISWA BIDIK MISI DI POLBENG," *JURNAL INOVTEK POLBENG- SERI INFORMATIKA*, pp. 87-94, 2016 .
- [27] Mustakim and G. Oktaviani, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," *Jurnal Sains, Teknologi dan Industri*, pp. 195 - 202 , 2016.
- [28] M. A. Banjarsari, I. Budiman and A. Farmadi, "Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4," *jurnal Ilmu Komputer (KLIK)* , pp. 50-64, 2015.
- [29] A. Rohman, "MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK PREDIKSI KELULUSAN MAHASISWA," *Jurnal Pandanaran Semarang* , pp. 1-9, 2015.
- [30] M. W. Amelia, A. S. M. Lumenta and A. Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes," *E-Journal Teknik Informatika* , pp. 1-7, 2017.
- [31] Fatmawati, "PERBANDINGAN ALGORITMA KLASIFIKASI DATA MININGMODEL C4.5 DAN NAIVE BAYES UNTUK PREDIKSI PENYAKIT DIABETES," *Jurnal Techno Nusa Mandiri* , pp. 50-59, 2016.
- [32] J. Han, Data Mining Concepts and Techniques, America: Elsevier, 2012.
- [33] Florin, Data Mining Concept And Techniques, India: Springer, 2011.
- [34] E. R. E. Sirait, "IMPLEMENTASI TEKNOLOGI BIG DATA DI LEMBAGA PEMERINTAHAN INDONESIA," *Jurnal Penelitian Pos dan Informatika* , pp. 113-136, 2016.
- [35] E. N. Wahyudi, " Teknik klasifikasi untuk melihat kecenderungan calon mahasiswa baru dalam memilih jenjang pendidikan program studi di perguruan tinggi," pp. 55-64, 2013.

- [36] H. Sulistiani and Y. T. Utami, "PENERAPAN ALGORITMA KLASIFIKASI SEBAGAI PENDUKUNG KEPUTUSAN PEMBERIAN BEASISWA MAHASISWA," *SNTI*, pp. 300-305, 2018.
- [37] Y. Farida and N. Ulinuha, "KLASIFIKASI MAHASISWA PENERIMA PROGRAM BEASISWA BIDIK MISI MENGGUNAKAN NAIVE BAYES," *SYSTEMIC*, pp. 17-22, 2018.
- [38] H. Rufaidha, F. Nhita and D. Triantoro, "PREDIKSI PENYAKIT MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR DAN ALGORITMA GENETIKA UNTUK DATA BERDIMENSI TINGGI," *e-Proceeding of Engineering*, pp. 3771-3777, 2016.
- [39] Risnawati, "ANALISIS KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA C.45," *Jurnal Mantik Penusa*, pp. 71-76, 2018.
- [40] L. Muflikhah, W. L. Yunita and M. T. Furqon, "Prediksi Nilai Mata Kuliah Mahasiswa Menggunakan Algoritma K-Apriori," *Jurnal Sisfo*, p. 157-172, 2017.
- [41] P. Irfansyah, "KAJIAN KOMPARASI PENERAPAN ALGORITMA DATA MINING (C4.5, BAYESIAN CLASSIFIER, DAN NEURAL NETWORK) DALAM MENENTUKAN PROMOSI JABATAN," *PROSIDING SEMINAR NASIONAL*, pp. 53-67, 2016.