

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
PERNYATAAN	iv
KATA PENGANTAR	vi
DAFTAR ISI	xi
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR SINGKATAN	xii
Intisari	xiii
<i>Abstract</i>	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	6
1.3. Tujuan Penelitian	7
1.4. Manfaat Penelitian	8
1.5. Batasan Masalah	8
1.6. Sistematika Penulisan	9
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	10
2.1. Tinjauan Pustaka	10
2.1.1. <i>Educational Data Mining</i>	10
2.1.2. Evaluasi Seleksi Fitur	11
2.1.3. Klasifikasi Dataset Kecil dan Faktor Pengaruh	12
2.1.4. Keakurasian Data	13
2.2. Dasar Teori	13
2.2.1. Faktor Performa Akademik Siswa	13
2.2.2. <i>Educational Data Mining</i>	14
2.2.3. <i>Data Preprocessing</i>	15
2.2.4. <i>Gain Ratio</i>	18

2.2.5. <i>Principal Component Analysis (PCA)</i>	20
2.2.6. <i>Classifier Subset Evaluator</i>	21
2.2.7. REPTree	21
2.2.8. Naive Bayes	23
BAB III METODE PENELITIAN	25
3.1. Sumber Data	25
3.2. Alat dan Bahan	29
3.3. Langkah Kerja	29
BAB IV HASIL DAN PEMBAHASAN	34
4.1. Pengujian Awal Klasifikasi	34
4.1.1. Pengujian Awal Naive Bayes	34
4.1.2. Pengujian Awal REPTree	37
4.2. Tahap <i>Preprocess</i> Lanjut	39
4.2.1. Pembersihan Data	39
4.2.2. Transformasi Data	42
4.2.3. <i>Loading Data</i>	45
4.3. Eksekusi disertai Seleksi Atribut	46
4.3.1. Seleksi Subset	46
4.3.1.1. Implementasi Naive Bayes	47
4.3.1.2. Implementasi REPTree	51
4.3.2. Seleksi Atribut	54
4.3.2.1. Implementasi Naive Bayes	55
4.3.2.2. Implementasi REPTree	57
4.3.3. Reduksi Dimensi Atribut	59
4.3.3.1. Implementasi Naive Bayes	60
4.3.3.2. Implementasi REPTree	63
4.4. Pembahasan	70
4.4.1. Peranan <i>preprocessing</i> pada Dataset Kecil	70
4.4.2. Peranan Jenis Validasi pada Dataset Kecil	71
4.4.3. Faktor yang Mempengaruhi Performa Siswa	73

BAB V KESIMPULAN DAN SARAN	78
4.1. Kesimpulan	78
4.2. Saran.....	80
DAFTAR PUSTAKA.....	81
LAMPIRAN.....	83

DAFTAR TABEL

Tabel 1.1	Hasil Survei PISA yang melibatkan 65 negara	1
Tabel 3.1	Indikator dan atribut yang tertera dalam angket siswa	27
Tabel 3.2	Translasi nilai menjadi nominal angka (Normalisasi)	32
Tabel 4.1	Perhitungan manual probabilitas dataset	36
Tabel 4.2	Sistematika penanganan nilai yang hilang di berbagai atribut	41
Tabel 4.3	Proses generalisasi data atribut Nilai Akhir	43
Tabel 4.4.	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada subset terbaik dengan Naive Bayes - CSE	50
Tabel 4.5	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada subset terbaik dengan REPTree - CSE	54
Tabel 4.6	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada atribut terbaik dengan Naive Bayes – GR	57
Tabel 4.7	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada atribut terbaik dengan REPTree – GR	59
Tabel 4.8	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada hasil reduksi atribut dengan Naive Bayes – PCA	63
Tabel 4.9	Ilustrasi translasi nilai nominal menjadi atribut biner	64
Tabel 4.10	Hasil validasi 10 fold cross dengan kombinasi 10 <i>random seed</i> pada hasil reduksi atribut dengan REPTree – PCA	70
Tabel 4.11	Ringkasan hasil dari seluruh penelitian	72
Tabel 4.12	Nilai faktor yang mempengaruhi baiknya performa belajar siswa	77
Tabel 4.13	Nilai faktor yang mempengaruhi buruknya performa belajar siswa	77

DAFTAR GAMBAR

Gambar 1.1	Presentase tingkat kebahagiaan siswa di sebagian negara dari 65 peserta	3
Gambar 2.1	Contoh penanganan nilai yang hilang	16
Gambar 2.2	Mekanisme <i>k-cross validation</i>	18
Gambar 3.1	Siswa kelas 9C SMP Muhammadiyah 2 Depok sedang mengisi angket	28
Gambar 3.2	Diagram pelaksanaan penelitian	30
Gambar 3.3	File ARFF yang dihasilkan	33
Gambar 4.1.	Hasil awal eksekusi Naive Bayes	35
Gambar 4.2	Perhitungan probabilitas dataset di aplikasi Weka	37
Gambar 4.3	Hasil awal eksekusi REPTree	38
Gambar 4.4	Bagan pelaksanaan preprocess lanjut	39
Gambar 4.5	Proses penanganan nilai yang kosong	41
Gambar 4.6	Proses pelaksanaan randomisasi dataset	42
Gambar 4.7	Skema pelaksanaan eksekusi penelitian	45
Gambar 4.8	Penggunaan Meta filter AttributeSelectedClassifier	46
Gambar 4.9	Hasil validasi holdout pada Subset Evaluator dengan Naive Bayes	49
Gambar 4.10	Subset atribut terbaik hasil ClassifierSubsetEval - REPTree	52
Gambar 4.11	Visualisasi tree hasil ClassifierSubsetEval – REPTree	52
Gambar 4.12	Hasil validasi holdout pada Subset Evaluator dengan REPTree	53
Gambar 4.13	Hasil validasi holdout pada Gain Ratio dengan Naive Bayes	56
Gambar 4.14	Hasil reduksi oleh PrincipalComponent dengan Naive Bayes	61
Gambar 4.15	Hasil validasi holdout pada PrincipalComponent dengan Naive Bayes	62
Gambar 4.16	Matriks kovarian dari keluaran Weka	64
Gambar 4.17	Nilai eigenvalues hasil keluaran Weka	66
Gambar 4.18	Hasil eigenvector yang berasal dari nilai eigenvalue	67
Gambar 4.19	Hasil reduksi oleh PrincipalComponent dengan REPTree	69
Gambar 4.20	Hasil validasi holdout pada PrincipalComponent dengan REPTree	70
Gambar 4.21	Tiga faktor pengaruh performa belajar siswa SMP Muhammadiyah 2 Depok, Sleman	75

DAFTAR SINGKATAN

C

CSE *Classifier Subset Evaluator*

CG *Conceptual Graph*

E

EDM *Educational Data Mining*

G

GolDar Golongan Darah siswa

GajiO Jumlah gaji orangtua siswa

GR *Gain Ratio*

I

ID Identitas nomor

J

Jen Jenis kelamin siswa

JumKel Jumlah anggota keluarga termasuk siswa

JamBel Jumlah jam belajar siswa dalam sehari

K

KTSP Kurikulum Tingkat Satuan Pendidikan

M

MapelFav Mata pelajaran yang digemari siswa

N

NB Naive Bayes

P

PddA Pendidikan Ayah

PddI Pendidikan Ibu

PkrA Pekerjaan Ayah

PkrI Pekerjaan Ibu

PCA *Principal Component Analysis*

R

REPTree *Reduced Error Pruning Tree*