

# Klasifikasi Soal Sejarah Tingkat SMA Berdasarkan Level Kognitif *Revised Bloom's Taxonomy* Menggunakan Metode *Stochastic Gradient Descent*

1<sup>st</sup> Amarila Zahratun Nisa  
Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

amarilanisa@student.telkomuniversity.ac.id

2<sup>nd</sup> Oktariani Nurul Pratiwi  
Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

onurulp@telkomuniversity.ac.id

3<sup>rd</sup> Riska Yanu Fa'rifah  
Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

riskayanu@telkomuniversity.ac.id

**Abstrak**— Pendidikan di tingkat formal membutuhkan aspek pengujian untuk memastikan ilmu pengetahuan diserap oleh siswa. Uji kompetensi terdiri dari berbagai macam tipe soal, salah satunya yaitu pilihan ganda. Soal yang dihimpun oleh guru dalam platform e-learning membutuhkan pemilahan sehingga dapat mengukur tingkat intelektual siswa dan mencakup keterampilan berpikir. Evaluasi dalam pemilahan soal didapati banyaknya soal yang belum tersaring sesuai dengan capaian kompetensinya. Pada penelitian ini, proses penyaringan yang melibatkan algoritma *Revised Bloom's Taxonomy* diimplementasikan secara otomatis dengan menggunakan metode *machine learning* yaitu *Stochastic Gradient Descent (SGD)* pada soal sejarah tingkat SMA. Dikarenakan terdapat imbalance pada dataset, penelitian ini akan membandingkan antara dataset yang menerapkan dan yang tidak menerapkan metode *oversampling SMOTE*. Hasil klasifikasi dari penelitian ini diperoleh melalui implementasi algoritma *SGD* dengan tools *python*. Pada dataset tanpa *SMOTE*, algoritma *SGD* memiliki skor rata-rata *K-Fold* yaitu 62%. Pada dataset dengan *SMOTE*, algoritma *SGD* memiliki skor rata-rata 93%. Adapun hasil akurasi dari *confusion matrix* menunjukkan algoritma *SGD* pada dataset tanpa *SMOTE* memiliki performa 62%, sedangkan algoritma *SGD* pada dataset dengan *SMOTE* mendapatkan performa 94%. Hasil yang diperoleh melalui proses *K-Fold Cross Validation* dan *confusion matrix* tersebut menunjukkan bahwa dataset dengan menggunakan *oversampling* memiliki hasil yang lebih baik dibanding dengan dataset tanpa menggunakan *oversampling*.

**Kata kunci**— klasifikasi pertanyaan, *RBT*, sejarah SMA, *SGD*, *oversampling*, *SMOTE*

## I. PENDAHULUAN

Pendidikan secara umum bertujuan untuk mengembangkan kepribadian diri manusia karena terdapat proses pengajaran ilmu pengetahuan dan keterampilan (*skills*). Sebagai upaya untuk menjamin siswa memahami materi yang diajarkan, guru mengadakan uji kompetensi dalam bentuk Penilaian Harian (PH), Penilaian Tengah Semester (PTS), dan Penilaian Akhir Semester (PAS). Ujian merupakan bentuk evaluasi proses belajar dalam rangka mengukur taraf pencapaian keahlian, karakter serta intelegensi siswa, sehingga menjadikan ujian sebagai tahapan penting dalam proses belajar mengajar [25].

*Higher order thinking skills* (HOTS) atau keterampilan berpikir tingkat tinggi merupakan bagian dari *Bloom's*

*Taxonomy* hasil revisi yang berupa kata kerja operasional yang terdiri dari *analyze* (C4), *evaluate* (C5) dan *create* (C6) yang dapat digunakan dalam penyusunan soal.

Guru menyusun bank soal salah satunya pada platform *e-Learning* untuk menghimpun pertanyaan yang akan diujikan untuk siswa. Penyusunan soal membutuhkan pemilahan hingga hanya soal tingkat tinggi yang dapat mengukur tingkat intelektual siswa dan mencakup keterampilan berpikir yang akan terpilih [19]. Mata pelajaran sejarah termasuk tipe pelajaran dengan banyak teori dan bacaan. Penggunaan *e-Learning* pada mata pelajaran ini penting karena mata pelajaran sejarah dapat melandasi pendidikan intelektual dan kedisiplinan siswa [18]. Meskipun begitu, penyusunan bank soal *e-Learning* terutama pada mata pelajaran sejarah memerlukan evaluasi ulang dikarenakan banyaknya soal yang belum tersaring sesuai dengan capaian kompetensinya. Proses penyaringan melibatkan algoritma identifikasi, salah satunya yang sering digunakan dalam peningkatan kualitas soal yaitu *Bloom's Taxonomy* [16].

Secara umum klasifikasi soal berdasarkan *Bloom's Taxonomy* dilakukan secara manual, namun cara tersebut dinilai relatif lama untuk mengolah data dalam skala besar [2]. Oleh karena itu, penulis mengajukan solusi untuk mengadakan klasifikasi soal secara otomatis dengan menggunakan salah satu metode *machine learning*. Metode *machine learning* diketahui dapat dipakai untuk memproses soal ujian dengan menggunakan *NLP* [14].

Metode yang akan digunakan pada penelitian ini yaitu *Stochastic Gradient Descent (SGD)*. *Stochastic Gradient Descent* menggunakan *gradien stochastic* yang meminimalkan fungsi kerugian yang dipilih dengan fungsi linear. Algoritma ini mendekati gradien yang benar dengan mempertimbangkan satu sampel pada suatu waktu, dan secara bersamaan memperbarui model berdasarkan gradien fungsi kerugian [23].

Berdasarkan uraian di atas, penelitian ini akan mengusulkan melakukan model dalam teknik klasifikasi pada *data mining* bank soal. Penelitian ini akan menghasilkan klasifikasi tipe soal berdasarkan tingkat kognitif, tingkat akurasi, dan prediksi guna penentuan klasifikasi soal sejarah SMA tipe C4 hingga C6.

II. KAJIAN TEORI

A. Revised Bloom's Taxonomy

Taksonomi Bloom (BT) [4] didesain ulang untuk mengatasi pendekatan dan teori pembelajaran baru yang berpotensi berguna dari akhir abad ke-20, seperti metakognisi [7] dan konstruktivisme [20]. Hasil akhirnya, RBT adalah kerangka dua dimensi yang memisahkan pengetahuan dan kognitif [1].

Dalam dimensi proses kognitif, enam kategori disajikan dalam format kata kerja dari kompleksitas kognitif rendah mengingat kompleksitas kognitif yang tinggi dari menciptakan [13]. Untuk membuat RBT lebih *user friendly* bagi dosen dan guru, "hirarki ketat" diabaikan untuk memungkinkan kategori tumpang tindih satu sama lain [13]. Dengan demikian, struktur hierarkis RBT tidak seketat BT.

Keenam tingkatan di dalam RBT yaitu [14]:

TABEL 1  
TINGKATAN KOGNITIF RBT

No.	Tingkat	Keterangan
1.	Remembering (C1)	Tingkat yang melibatkan kemampuan untuk mengingat dan menyebutkan informasi atau pengetahuan yang diperoleh.
2.	Understanding (C2)	Tingkat yang melibatkan kemampuan untuk memahami instruksi dan menceritakan konsep atau ide yang didapatkan, baik secara lisan, tertulis, maupun media seperti diagram maupun grafik.
3.	Applying (C3)	Tingkat yang melibatkan kemampuan untuk menerapkan ide atau konsep dalam kondisi tertentu.
4.	Analyzing (C4)	Tingkat yang melibatkan kemampuan untuk memecah konsep menjadi beberapa komponen dan menghubungkan satu sama lain untuk mendapatkan pemahaman baru secara keseluruhan.
5.	Evaluating (C5)	Tingkat yang melibatkan kemampuan untuk menilai sesuatu berdasarkan kriteria atau standar tertentu.
6.	Creating (C6)	Tingkat yang melibatkan kemampuan untuk mengintegrasikan berbagai elemen menjadi sesuatu yang baru dan orisinal.

B. Klasifikasi Pertanyaan (*Question Classification*)

Klasifikasi pertanyaan adalah proses dimana sistem menganalisis pertanyaan dan label [15]. Pertanyaan tersebut ditujukan berdasarkan jenis jawaban yang diharapkan. Misalnya, pertanyaan "Siapa nama anak Pangeran Diponegoro?" mengharapkan nama seseorang sebagai jawaban. Tujuan dari sistem klasifikasi pertanyaan adalah untuk mempelajari pemetaan dari pertanyaan untuk menjawab jenis dari pertanyaan tersebut [15]. Klasifikasi pertanyaan sangat penting untuk menjawab pertanyaan.

C. Penambangan Teks (*Text Mining*)

*Text mining* adalah penemuan dalam dunia komputer baru yang secara otomatis mengekstraksi informasi dari sumber daya tertulis yang berbeda [10].

*Text mining* dalam prakteknya erat dengan data teks yang paling disamakan dan tidak terstruktur, sehingga memiliki kontak yang sangat dekat dengan bidang lain, seperti pengambilan informasi, penyaringan informasi, ringkasan otomatis, *text clustering*, *text classification*, *natural language processing*, *artificial intelligence*, *machine learning*, *pattern recognition*, statistik, visualisasi dan sebagainya [24].

D. Text Preprocessing

*Text preprocessing* merupakan metode yang sangat penting dalam teknik dan aplikasi penambangan teks. Ini adalah langkah pertama dalam proses penambangan teks [26]. *Text Preprocessing* yang akan digunakan terbagi menjadi beberapa tahap [17], yaitu *case folding*, *tokenizing*, *stopwords removal*, *stemming*, dan *term weighting*.

E. Term Frequency – Inverse Document Frequency (TF-IDF)

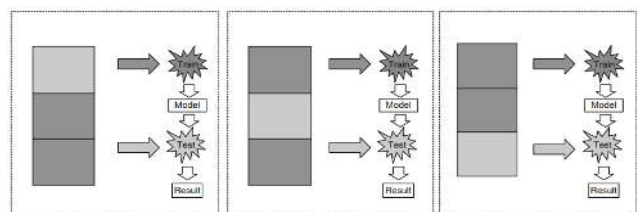
TF-IDF adalah bentuk representasi dokumen yang paling mendasar dan memiliki sejarah terpanjang di antara metode representasi yang umum digunakan [22]. TF-IDF memberikan asumsi bahwa jika sebuah kata itu penting untuk sebuah dokumen, kata tersebut harus berulang kali muncul dalam dokumen itu sedangkan jarang muncul dalam dokumen lain [12]. TF dikaitkan dengan asumsi sebelumnya sedangkan IDF dikaitkan dengan asumsi yang terakhir. TF-IDF didefinisikan dalam persamaan berikut.

$$TF - IDF_{ij} = tf_{ij} \times \log \left( \frac{N}{df_{i+1}} \right) \tag{1}$$

Parameter  $tf_{ij}$  didefinisikan sebagai berapa kali kata  $i$  muncul dalam dokumen  $j$ ; Semakin besar nilainya, semakin penting kata itu. Parameter  $df_i$  adalah jumlah dokumen di mana kata  $i$  muncul setidaknya sekali. Semakin besar nilainya, semakin umum kata itu. Jika kata  $i$  dapat dianggap penting untuk dokumen  $j$ , kata tersebut harus memiliki TF besar ( $tf_{ij}$ ) dan DF kecil ( $df_i$ ).

F. Cross-Validation

*Cross-Validation* adalah metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua segmen; satu digunakan untuk mempelajari atau melatih model dan yang lainnya digunakan untuk memvalidasi model [21]. Dalam *K-Fold cross-validation*, data pertama kali dipartisi menjadi segmen atau lipatan berukuran sama (atau hampir sama). Selanjutnya, iterasi  $k$  pelatihan dan validasi dilakukan sedemikian rupa sehingga dalam setiap iterasi lipatan data yang berbeda diadakan untuk validasi, sedangkan sisa  $k - 1$  lipatan digunakan untuk pembelajaran. Gambar 1 menunjukkan contoh dengan  $k = 3$ .



GAMBAR 1  
K-3 K-FOLD CROSS VALIDATION

G. Evaluasi Performansi

*Confusion matrix* biasanya digunakan dalam pembelajaran mesin untuk mengevaluasi atau memvisualisasikan perilaku model dalam konteks *supervised classification* [11]. *Confusion matrix* berbentuk persegi di mana setiap baris mewakili kelas aktual dari instance dan kolom kelas prediksi [5].

TABEL 2  
STRUKTUR *CONFUSION MATRIX*

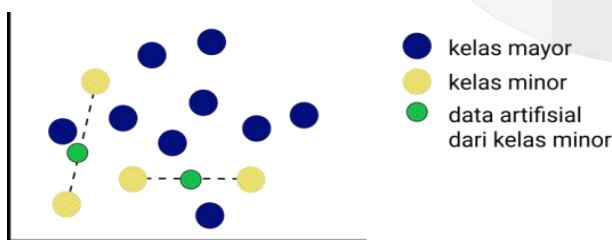
	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

*False Negative* (FN) merupakan data positif yang diprediksi sebagai data negative, sedangkan *False Positive* (FP) merupakan data negatif yang diprediksi sebagai data positif. Adapun *True Negative* (TN) merupakan data negatif yang diprediksi benar, sedangkan *True Positive* (TP) merupakan data positif yang diprediksi benar.

H. *Synthetic Minority-Over Sampling Technique (SMOTE)*

Teknik *preprocessing* SMOTE menjadi pelopor bagi komunitas penelitian dalam klasifikasi *imbalance learning*. Karena popularitas dan pengaruhnya, SMOTE dianggap sebagai salah satu algoritma *preprocessing/sampling* paling berpengaruh dalam *machine learning* dan *data mining* [9]. Keberhasilan teknik ini tentunya terkait dengan empat faktor utama: (1) kesederhanaan dan kemudahan implementasi dalam aplikasi apa pun, (2) jaminan peningkatan kinerja, (3) generalitas dengan semua jenis algoritma pembelajaran dan (4) kemampuan beradaptasi dan berekstensi dalam pendekatan baru.

SMOTE bekerja dengan menggunakan teknik pembangkitan amatan buatan pada kelas minor berdasarkan konsep *k*-tetangga terdekat. Nilai *k*-tetangga terdekat dipilih secara acak tergantung dari diperlukan atau tidaknya pengambilan *oversampling* [6]. Jumlah *k*-tetangga terdekat ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya [8].



GAMBAR 2  
METODE SMOTE

I. *SGD (Stochastic Gradient Descent)*

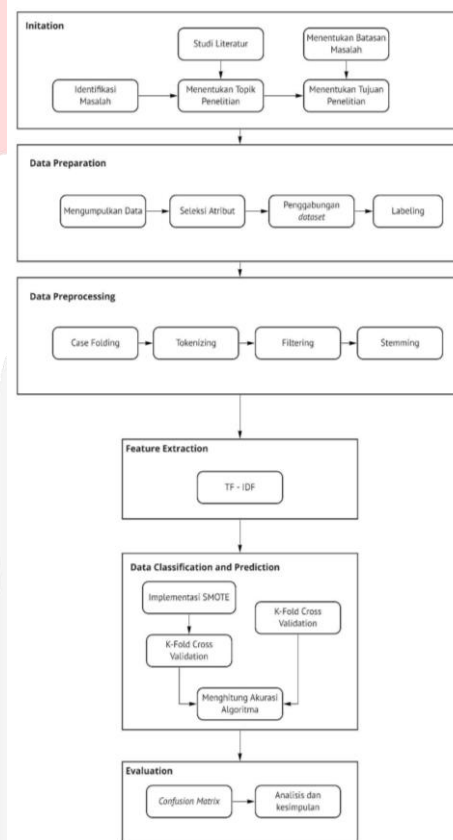
*Stochastic Gradient Descent* (SGD) merupakan metode *Gradient Descent* yang melakukan update parameter untuk setiap data pelatihan  $x^{(i)}$  serta label  $y^{(i)}$  dan memiliki persamaan dasar sebagai berikut:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}y^{(i)}) \quad (2)$$

*Gradient Descent* adalah cara untuk meminimalkan fungsi tujuan  $J(\theta)$ .  $\theta \in \mathbb{R}^d$  merupakan parameter model.  $\eta$  merupakan kecepatan belajar.  $\nabla_{\theta} J(\theta)$  merupakan gradien fungsi tujuan terhadap parameter. Parameter diperbarui dalam arah gradien yang berlawanan. Berdasarkan fungsinya, *Stochastic Gradient Descent* (SGD) memiliki kelebihan yaitu ketersediaan dalam penggunaan sumber belajar *online* yang tidak ada pada *Gradient Descent* sebelumnya, sedangkan kelemahan berupa tingginya pembaruan pada varian [23].

III. METODE

Penelitian ini menggunakan sistematika penyelesaian masalah melalui beberapa tahap yaitu inialisasi, pengolahan data, kemudian mengekstraksi hasil dan kesimpulan. Diagram sistematika penyelesaian masalah dapat dilihat pada gambar di bawah ini.



GAMBAR 3  
DIAGRAM SISTEMATIKA PENYELESAIAN MASALAH

A. *Initiation*

Proses *initiation* terdiri dari identifikasi masalah dan studi literatur yang menjadi awal dari penentuan topik penulisan, kemudian bersama dengan penentuan batasan masalah akan ditentukan tujuan penelitian ini.

B. *Data Preparation*

Proses kedua yaitu *data preparation* yang terdiri dari pengumpulan data, seleksi atribut data, penggabungan *dataset*, dan *labelling data*. Data yang terhimpun dari internet adalah kumpulan soal Latihan, Penilaian Harian (PH), Penilaian Tengah Semester (PTS), Penilaian Akhir Semester (PAS), dan Ujian Nasional dari kelas X, XI, dan XII mata pelajaran Sejarah SMA. Soal-soal tersebut diunduh dengan

bentuk PDF dan bertipe pilihan ganda. Selanjutnya dipilih hanya pertanyaan saja untuk diolah. Total soal yang didapatkan 58 butir soal C4-C6. Berikutnya, soal-soal tersebut diberi label sesuai kategori RBT, dilakukan kategorisasi tingkat kesukaran menurut pembuat soal dan dihimpun dalam kesatuan *dataset* berbentuk *spreadsheet* untuk diproses menuju tahap selanjutnya. Pada penelitian ini, *labelling data* dilakukan secara manual.

C. *Data Preprocessing*

Pada tahap ini, data akan diolah melalui beberapa langkah. Proses yang dilakukan yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming* pada *dataset*. Pelaksanaannya menggunakan *tools* Python dengan melibatkan library NLTK dan Sastrawi.

D. *Feature Extraction*

Tahap selanjutnya yaitu *feature extraction* dengan menggunakan metode TF-IDF. Tujuan dari pembobotan supaya algoritma dapat memproses teks pada data secara langsung sehingga hasilnya akan dikonversi menjadi bentuk numerik. Proses pertama pada tahap ini melibatkan perhitungan soal (*Term Frequency*) dan pembobotan kata (*Inverse Document Frequency*). Proses kedua kemudian seluruh kata dalam dokumen akan menghasilkan bobot untuk perhitungan klasifikasi menggunakan *classifier*.

E. *Data Classification and Prediction*

Proses kelima berupa *data classification and prediction* yang terdiri dari implementasi SMOTE, *K-Fold Cross Validation*, dan menghitung akurasi algoritma. *K-Fold Cross Validation* diterapkan sebanyak dua kali pada *dataset* tanpa dan dengan pengimplementasian SMOTE.

F. *Evaluation*

Tahap terakhir yaitu perhitungan *confusion matrix*. Penerapan *confusion matrix* dilakukan untuk membandingkan hasil penggunaan algoritma dengan dan tanpa penggunaan *oversampling*. Hasil akurasi diperoleh kemudian dianalisis hingga membuahkan kesimpulan hasil penelitian.

IV. HASIL DAN PEMBAHASAN

Sumber data terdiri dari 10 *file* bank soal dengan jumlah soal berbeda tiap *file*. Masing-masing terdiri dari gabungan soal sejarah kelas X, XI, dan XII mata pelajaran Sejarah SMA. Dari keseluruhan soal dipilih 304 soal dengan tingkat RBT C1-C6 dan dipisah kembali untuk RBT C4-C6 hingga menghasilkan 58 soal. Kemudian peneliti akan menggabungkan kelima kolom soal menjadi satu kolom soal. Setelah itu, peneliti akan menghilangkan bagian pilihan ganda pada masing-masing soal sehingga hanya tersisa pertanyaannya saja. Selanjutnya, *dataset* dipilih untuk menghilangkan redundansi data dan memperbaiki kalimat yang salah ejaan. Kemudian, *dataset* di-export ke dalam file dengan format *csv* atau *excel* untuk memasuki tahap *labelling*. Tahap selanjutnya, *dataset* diberi label secara manual untuk mendapatkan kolom label berdasarkan level kognitif RBT seperti berikut.

TABEL 3  
CONTOH SOAL DAN LABEL

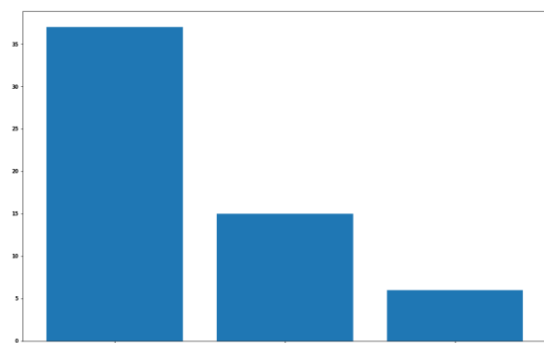
Soal	Label
Salah satu penyebab kacaunya kondisi perekonomian Indonesia pada masa awal kemerdekaan karena kas negara kosong. Upaya pemerintah Republik Indonesia mengisi kas negara yang kosong pada awal Kemerdekaan adalah	Analyzing/ C4
Gerakan Non Blok merupakan organisasi negara-negara berkembang yang dibentuk berdasarkan Konferensi Beograd pada tahun 1961. Organisasi yang digagas antara lain oleh Bung Karno tersebut bertujuan	Evaluating/ C5
"Kami bangsa Indonesia dengan ini menyatakan kemerdekaan Indonesia". Makna penggalan pernyataan Proklamasi Kemerdekaan Indonesia tersebut bagi bangsa Indonesia adalah	Creating/ C6

Setelah dilakukan *preprocessing* terhadap *dataset*, tahap selanjutnya yaitu pembobotan kata menggunakan TF-IDF. Tabel berikut menunjukkan hasil akhir dari *data preprocessing* (kolom kiri) dan TF-IDF (kolom kanan).

TABEL 4  
PEMBOBOTAN TF IDF SETELAH STEMMING

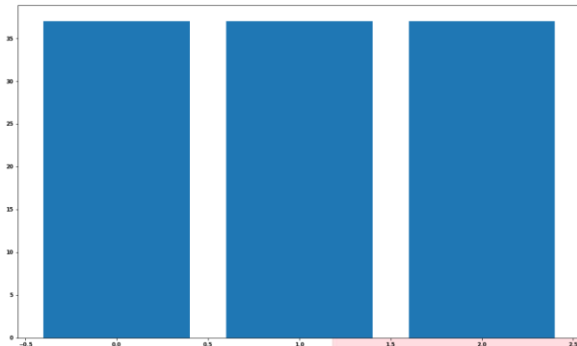
Stemming	TF-IDF
['gera', 'non', 'blok', 'organisasi', 'negeranegara', 'kembang', 'bentuk', 'dasar', 'konferensi', 'beograd', '1961', 'organisasi', 'gagas', 'karno', 'tuju']	'gera': 0.18683626625985425 'non': 0.2855830402351563 'blok': 0.2855830402351563 'organisasi': 0.41389331566186865 'negeranegara': 0.25896315254118285 'kembang': 0.2134561539538277 'bentuk': 0.20694665783093433 'dasar': 0.21855153342075778 'konferensi': 0.2855830402351563 'beograd': 0.2855830402351563 'gagas': 0.24007604164780108 'karno': 0.25896315254118285 'tuju': 0.1971348753245809

Tahap berikutnya yaitu *data classification and prediction*. Setelah diketahui terdapat *imbalance* pada *data*, *oversampling* diterapkan untuk menangani hal tersebut.



GAMBAR 4  
DATASET SEBELUM OVERSAMPLING

Berdasarkan Gambar 4, label C1 memiliki kuantitas terbanyak dengan jumlah 37 butir, label C5 sebanyak 15 butir, dan label C6 sebanyak 6 butir soal. Atas dasar ketidakseimbangan data yang dapat mempengaruhi hasil akhir, perlu diterapkan *oversampling* sebagai penanganan. SMOTE digunakan sebagai tipe *oversampling* pada penelitian ini. Setelah *dataset* diterapkan metode *oversampling*, maka hasilnya seperti tertera pada Gambar 5.



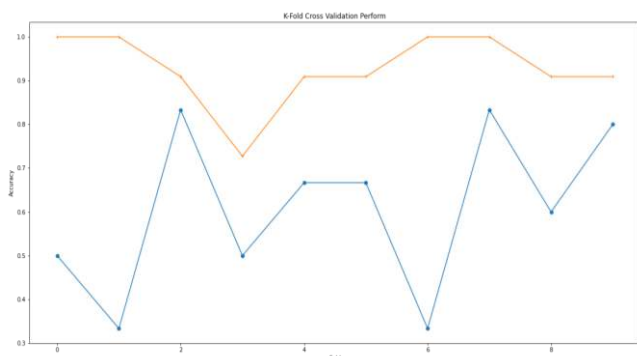
GAMBAR 5 DATASET SETELAH OVERSAMPLING

Total keseluruhan data setelah diterapkan SMOTE adalah sebanyak 111 butir soal. *Dataset* dikatakan telah seimbang dan siap untuk melanjutkan proses. Selanjutnya, *dataset* diuji dengan algoritma SGD dan *K-Fold Cross Validation*. Pada penelitian ini, nilai *k* yang digunakan yaitu *10-fold*. Tabel berikut menunjukkan hasil perbandingan sebelum dan sesudah *oversampling*.

TABEL 5 PERBANDINGAN TOTAL SKOR K-FOLD TERHADAP DATASET

Bentuk Implementasi	Hasil Skor
SGD Sebelum SMOTE	62%
SGD Sesudah SMOTE	93%

Hasil perbandingan skor berupa grafik yang mendokumentasikan akurasi sebanyak 10 *fold* ditunjukkan pada Gambar 6.



GAMBAR 6 PERBANDINGAN HASIL K-FOLD CROSS VALIDATION ALGORITMA SGD

Berdasarkan Gambar 6, algoritma dengan SMOTE (ditunjukkan dengan garis jingga) memiliki hasil yang lebih stabil dan akurasi lebih tinggi dibanding dengan algoritma tanpa SMOTE (garis biru). Proses selanjutnya yaitu *evaluation* menggunakan *confusion matrix*. Tabel 6 menampilkan hasil *confusion matrix* dari algoritma SGD tanpa SMOTE.

TABEL 6. CONFUSION MATRIX DARI SGD TANPA SMOTE

True / Predicted	C4	C5	C6	All
C4	31	4	2	37
C5	11	3	1	15
C6	6	0	0	6

Berdasarkan Tabel 6, model *classifier* SGD terhadap *dataset* tanpa SMOTE memprediksi label C4 dengan benar (TP) sebanyak 31, label C5 sebanyak 3 dan label C6 sebanyak 0. Label C4 diketahui paling banyak diprediksi dengan benar oleh model. Selanjutnya Tabel 7 merupakan hasil *confusion matrix* dari algoritma SGD dengan SMOTE.

TABEL 7 CONFUSION MATRIX DARI SGD TANPA SMOTE

True / Predicted	C4	C5	C6	All
C4	31	4	2	37
C4	33	4	0	37
C5	3	34	0	37
All	36	38	37	111

Berdasarkan Tabel 7, model *classifier* SGD terhadap *dataset* dengan SMOTE memprediksi label C4 dengan benar (TP) sebanyak 33, label C5 sebanyak 34, dan label C6 sebanyak 37. Label C6 paling banyak diprediksi dengan benar oleh model. Selanjutnya peneliti melakukan evaluasi performansi dari keempat *confusion matrix*. Hasil evaluasi performansi dapat dilihat pada Tabel 8 di bawah ini.

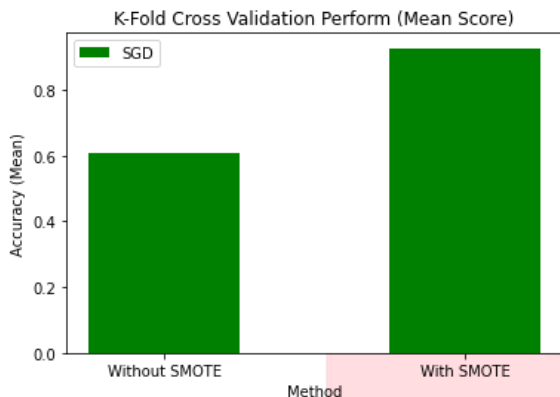
TABEL 8 HASIL EVALUASI PERFORMANSI

Metode	Precision	Recall	F1-Measure	Accuracy
SGD tanpa SMOTE	37%	38%	36%	62%
SGD dengan SMOTE	94%	94%	94%	94%

Pada Tabel 8, nilai *precision*, *recall*, dan *f1-measure* diambil berdasarkan nilai *weighted average* pada masing-masing *confusion matrix*. Nilai akurasi yang diperoleh oleh *classifier* terhadap *dataset* tanpa SMOTE yaitu 62% dari keseluruhan 37 *data testing*. Nilai akurasi yang diperoleh oleh *classifier* terhadap *dataset* dengan SMOTE yaitu 94% dari keseluruhan 111 *data testing*.

Perolehan nilai *accuracy* berarti hasil akurat bagaimana model bekerja dalam mengklasifikasikan soal dengan benar [3], dikatakan sudah baik dengan perolehan skor yang mencapai 94%. *Precision* berarti hasil akurat dari data terhadap hasil prediksi yang dihasilkan sendiri oleh model [3], dikatakan cukup baik dengan perolehan skor 94%. *Recall* berarti keberhasilan suatu model untuk menemukan kembali

informasi [3], dikatakan sudah baik dengan perolehan skor 94%. *F1-score* berarti perbandingan hasil bobot rata-rata *precision* dan *recall* [3], dikatakan sudah baik dengan perolehan skor 94%. Gambar 7 menunjukkan perbandingan hasil akurasi berdasarkan Tabel 8.



GAMBAR 7  
PERBANDINGAN RATA-RATA NILAI K-FOLD CROSS  
VALIDATION

Berdasarkan perbandingan tersebut, algoritma SGD dapat dipakai sebagai model untuk melakukan klasifikasi dan prediksi soal berdasarkan tingkat kognitif RBT. Namun, jumlah soal per label dapat lebih dipertimbangkan untuk meminimalisir adanya *imbalance data*. Penelitian ini menyimpulkan bahwa *dataset* yang lebih seimbang menghasilkan akurasi yang lebih tinggi daripada *dataset* yang tidak seimbang. Metode SMOTE juga dapat mempengaruhi hasil akurasi pada algoritma SGD.

## V. KESIMPULAN

Penelitian ini membahas mengenai klasifikasi soal sejarah SMA berdasarkan level kognitif RBT menggunakan implementasi algoritma SGD. Klasifikasi dilakukan menggunakan bank soal sejarah SMA yang dihimpun dari berbagai *website* dan berjumlah total 304 butir. Soal kemudian dikategorikan sesuai level kognitif C1-C6. Berdasarkan penyortiran tersebut, didapatkan 58 butir soal dengan level C4-C6 yang diproses pada penelitian ini. Tahapan penelitian terdiri dari *initiation*, *data preparation*, *data preprocessing*, *feature extraction*, *data classification and prediction*, dan *evaluation*. Hasil klasifikasi dari penelitian ini diperoleh melalui implementasi algoritma SGD dengan tools *python*. Pada *dataset* tanpa SMOTE, algoritma SGD memiliki skor rata-rata *K-Fold* yaitu 62%. Pada *dataset* dengan SMOTE, algoritma SGD memiliki skor rata-rata 93%. Adapun hasil akurasi dari *confusion matrix* menunjukkan algoritma SGD pada *dataset* tanpa SMOTE memiliki performa 62%, sedangkan algoritma SGD pada *dataset* dengan SMOTE mendapatkan performa 94%. Peneliti menyarankan untuk penelitian serupa di masa mendatang agar menggunakan metode lainnya baik algoritma maupun metode dalam penanganan *imbalance data*. Hal ini bertujuan guna memperoleh metode yang terbaik bagi penanganan metode klasifikasi.

## REFERENSI

[1] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (Eds.). (2001). A taxonomy for learning, teaching, and assessing:

A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York, N.Y.: Longman

[2] Aninditya, A., Hasibuan, M. A., dan Sutoyo, E. (2019). Text mining approach using TF-IDF and naive bayes for classification of exam questions based on cognitive level of bloom's taxonomy. Proceedings - 2019 IEEE International Conference on Internet of Things and Intelligence System, IoTaIS 2019.

[3] Anggreany, M. S. (2020, November 1). *Confusion Matrix*. Retrieved from BINA NUSANTARA School of Computer Science: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/>

[4] Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, 20(24), 1.

[5] Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3), 429-450.

[6] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.

[7] Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.

[8] Fuadin, D. N. (2017). *Deteksi Botnet Menggunakan Naive Bayes Classifier dengan SMOTE dan Metode BFS* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[9] Garcia, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98, 1-29.

[10] Hearst, M. (2003). What is text mining. SIMS, UC Berkeley, 5.

[11] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

[12] Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Information Sciences*, 477, 15-29.

[13] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.

[14] Kusuma, S. F., Siahaan, D., dan Yuhana, U. L. (2016). Automatic Indonesia's questions classification based on bloom's taxonomy using Natural Language Processing a preliminary study. 2015 International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings.

[15] Metzler, D., & Croft, W. B. (2005). Analysis of statistical question classification for fact-based questions. *Information Retrieval*, 8(3), 481-504.

[16] Mohammed, M., Omar, N. (2020). Question classification based on Bloom's Taxonomy using enhanced TF-IDF. *International Journal on Advanced Science, Engineering and Information Technology*.

[17] Najjichah, H., Syukur, A., & Subagyo, H. (2019). PENGARUH TEXT PREPROCESSING DAN KOMBINASINYA PADA PERINGKAS DOKUMEN OTOMATIS TEKS BERBAHASA INDONESIA. *Jurnal Cyberku*, 15(1), 1-11.

[18] Nurholif, D. M., Suartama, I. K., & Sukmana, A. I. W. I. Y. (2021). Belajar Sejarah Dengan E-Learning Berbasis Discovery Learning. *Mimbar Ilmu*, 26(3).

[19] Pantiwati, Y., & Permana, F. H. (2017). Analisis butir soal oleh mahasiswa S1 pendidikan biologi universitas muhammadiyah malang (UMM) berdasarkan PISA dan taksonomi bloom revisi. Seminar Nasional Pendidikan Berkemajuan Dan Menggembirakan (The Progressive & Fun Education Seminar) Ke-2.

[20] Piaget, J. (1980). *Six psychological studies*. Harvester Press.

[21] Refaeilzadeh P., Tang L., Liu H. (2016) Cross-Validation. In: Liu L., Özsu M. (eds) Encyclopedia of Database Systems. Springer, New York, NY.

[22] Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of documentation*.

[23] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

[24] Sangodiah, A., Ahmad, R., dan Ahmad, W. F. W. (2014). A review in feature extraction approach in question classification using Support Vector Machine. Proceedings - 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014, November, 536-541

[25] Sidiq, A. B., & Kurniadi, D. (2021). Perancangan Sistem Informasi Ujian Online Berbasis Web pada SMK N 1 Solok. *VoteTEKNIKA: Jurnal Vocational Teknik Elektronika dan Informatika*, 9(2), 44-53.

- [26] Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.

