

BAB II TINJAUAN PUSTAKA

Berikut adalah beberapa jurnal dan penelitian mengenai *Support Vector Machine* yang dijadikan sebagai acuan penelitian ini:

Penelitian yang dilakukan oleh Nugroho, et al., (2003) tentang klasifikasi *Support Vector Machine* aplikasinya dalam bioinformatika, eksperimen ini dilakukan oleh group Terrence S. Furey, dengan tujuan memakai SVM untuk mengklasifikasikan apakah suatu pasien terkena penyakit kanker atau tidak, berdasarkan hasil analisa microarray terhadap sel pasien tersebut. Secara umum, SVM menunjukkan hasil yang lebih baik daripada perceptron. Walaupun demikian, dikarenakan jumlah sampel yang relatif sedikit, hasil eksperimen itu belum dapat memberikan kesimpulan final bahwa SVM superior terhadap perceptron.

Penelitian yang dilakukan oleh Suryanto dan Purnami (2015) tentang perbandingan *Reduced Support Vector Machine* dan *Smooth Support Vector Machine* untuk Klasifikasi *Large Data* pada SVM memanfaatkan optimasi dengan *quadratic programming*, sehingga untuk data berdimensi tinggi dan berjumlah besar, SVM menjadi kurang efisien. Untuk mengatasi hal tersebut, dikembangkan *Smooth Support Vector Machine* (SSVM). Pada jumlah data yang besar SSVM juga tidak efisien kemudian dikembangkan *Reduced Support Vector Machine* (RSVM). Hasil penelitian menunjukkan pada jumlah data yang relatif kecil (kurang dari 1000) metode SSVM dan RSVM memberikan performansi yang sama, tetapi pada data yang relatif besar (lebih dari 1000) RSVM memberikan performansi yang lebih baik daripada SSVM.

Penelitian yang dilakukan oleh Barro, et al., (2013) tentang penerapan *Synthetic Minority Oversampling Technique* (SMOTE) terhadap data tidak seimbang pada pembuatan model komposisi jamu. Dengan didapatkan kesimpulan yaitu pada ukuran kebaikan model ditunjukkan oleh nilai R^2 Nagelkerke. Nilai R^2 Nagelkerke yang dihasilkan model dengan SMOTE lebih rendah 3.2% dibandingkan dengan R^2 Nagelkerke yang dihasilkan model tanpa SMOTE. Meskipun demikian, model dengan SMOTE lebih akurat karena nilai AUC yang

dihasilkan lebih tinggi daripada model tanpa SMOTE. Model dengan SMOTE memiliki nilai AUC sebesar 0.976 sedangkan model tanpa SMOTE memiliki nilai AUC sebesar 0.908. Hasil tersebut menunjukkan bahwa SMOTE dapat menaikkan tingkat akurasi model pada data tidak seimbang.

Penelitian yang dilakukan oleh Siringoringo (2018) dengan judul klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan *K-Nearest Neighbor*, Ketidakseimbangan data memberikan dampak yang buruk pada hasil klasifikasi dimana kelas minoritas sering disalah klasifikasikan sebagai kelas mayoritas. *k-Nearest Neighbor* merupakan salah satu metode klasifikasi yang sangat populer dan sederhana tetapi, tidak dilengkapi dengan kemampuan untuk bekerja pada dataset tidak seimbang. Pada penelitian ini, Synthetic Minority Over-sampling Technique (SMOTE) diterapkan untuk menyelesaikan masalah ketidak seimbangan kelas pada dataset Credit Card Fraud. Dengan menerapkan skema evaluasi 10-cross fold validation diperoleh hasil bahwa SMOTE meningkatkan rata-rata G-Mean dari 53,4% ke 81,0% dan rata-rata F-Measure dari 38,7 ke 81,8%.

Penelitian yang dilakukan oleh Li dan Hu (2019) dengan judul *Research on unbalanced training samples based on SMOTE algorithm*. Dalam klasifikasi, memiliki distribusi data *training* yang tidak seimbang merupakan masalah serius. Hal tersebut akan menyebabkan bias ke sampel kelas mayoritas dan tidak dapat mempelajari karakteristik data sampel pada kelas minoritas dengan baik, sehingga sulit untuk mengidentifikasi sampel kelas minoritas. Untuk meningkatkan kinerja jaringan saraf dalam pada sampel kelas minoritas, makalah ini mengusulkan skema baru berdasarkan algoritma SMOTE, yang menghasilkan sampel kelas minoritas baru untuk menyeimbangkan dataset pelatihan. Dibandingkan dengan pengambilan sampel dengan cara lama biasanya menggunakan klasifikasi gambar seperti menyalin gambar secara sederhana, membalik gambar, menyesuaikan warna dll., Skema penelitian ini mengeneralisasikan sampel kelas minoritas baru, yang meningkatkan kualitas sampel kelas minoritas dan menghindari *over-fitting* pada saat dalam klasifikasi. Alur kerja *preprocessing* gambar dan operasi pengambilan sampel berlebihan SMOTE diperkenalkan dalam makalah ini. Basis pemilihan model jaringan dan indikator model komprehensif juga disajikan. Eksperimen komparatif dilakukan dengan menguji model yang dilatih masing-masing pada

dataset tidak seimbang dan dataset seimbang setelah operasi pengambilan sampel berlebihan SMOTE. Hasil menunjukkan bahwa skema yang diusulkan layak dan efektif untuk meningkatkan kemampuan jaringan saraf mengidentifikasi sampel kelas minoritas dalam tugas klasifikasi.

Penelitian oleh Pernes, et al., (2019) dengan judul *Directional Support Vector Machines*, tentang Beberapa fenomena diwakili oleh data *directional—angular* atau periodik; dari referensi waktu pada kalender ke koordinat geografis. Nilai-nilai ini biasanya direpresentasikan sebagai nilai nyata yang dibatasi pada rentang tertentu (seperti $[0, 2\pi)$), menyembunyikan sifat sebenarnya dari informasi ini. Untuk menangani variabel-variabel ini dengan benar dalam tugas-tugas klasifikasi *supervised*, alternatif untuk *classifier naif Bayes* dan regresi logistik diusulkan di penelitian sebelumnya. Dalam karya ini, peneliti mengusulkan mesin vektor dukungan *directional-aware*. Kami membahas beberapa realisasi dari model yang diusulkan, mempelajari kernel mereka dan ekspresifnya. Akhirnya, kami memvalidasi kinerja Mesin Vector Support yang diusulkan (SVMs) terhadap *directional naes Bayes* dan regresi logistik *directional* dengan data nyata, memperoleh hasil kompetitif.

Penelitian yang dilakukan oleh Munoz, et al., (2019) dengan judul *Support Vector Machine*, pada penelitian ini akan mendeskripsikan inti dari pembelajaran *support vector*. Ini termasuk gagasan belajar dari data melalui minimalisasi risiko dan gagasan kernel sebagai cara yang efisien untuk menyandikan informasi sebelumnya. peneliti memberikan tinjauan umum atas masalah optimasi yang timbul dari klasifikasi *support vector*, regresi, estimasi kuantil dan kombinasi informasi. Selain itu, peneliti juga membahas beberapa ekstensi seperti *kernel Principal Component Analysis*, *Independent Component Analysis* dan *Manifold Learning*.

Penelitian yang dilakukan oleh Amalia Dwi Nurfadzilah (2019) tentang klasifikasi nilai mata kuliah Mahasiswa Statistika UII angkatan 2015 Berdasarkan rekam jejak akademik meliputi persentase keterlambatan kehadiran kuliah, jenis mata kuliah yang diambil, jam mata kuliah, jadwal hari mata kuliah dan jumlah

SKS menggunakan *Support Vector Machine*. Serta SMOTE (*Synthetic Minority Oversampling Technique*) pada proses untuk meningkatkan performa klasifikasi SVM. Dengan mendapatkan tingkat akurasi 59% dengan Cost yaitu 10 dan gamma 100 didapatkan bahwa yang masuk kedalam kelas “Amat baik” yaitu sebanyak 363, “baik” sebanyak 102, “cukup” sebanyak 4, “kurang” hanya 2, “sangat kurang” hanya satu, serta “tidak memenuhi” sebanyak 66.

Tabel 2.1 Tabel Sumber Jurnal

Judul Penelitian	Peneliti, tahun	Metode	Tujuan
<i>Support Vector Machine</i> Teori dan Aplikasinya Dalam Bioinformatika	Nugroho et al., (2003)	<i>Support Vector Machine</i>	tujuan memakai SVM untuk mengklasifikan apakah suatu pasien terkena penyakit kanker atau tidak, berdasarkan hasil analisa microarray terhadap sel pasien tersebut.
<i>Reduced Support Vector Machine</i> dan <i>Smooth Support Vector Machine</i> untuk Klasifikasi <i>Large Data</i>	Suryanto dan Purnami (2015)	<i>Support Vector Machine</i> , <i>Smooth Support Vector Machine</i> , dan <i>Reduced Support Vector Machine</i>	Klasifikasi <i>Large Data</i> pada SVM dengan memanfaatkan optimasi dengan <i>quadratic programming</i> untuk data berdimensi tinggi dan berjumlah besar.
Penerapan <i>Synthetic Minority Oversampling Technique</i> (SMOTE) terhadap Data	Barro, et al., (2013)	SMOTE (<i>Synthetic Minority Oversampling Technique</i>)	Penelitian ini bertujuan membandingkan model komposisi jamu yang berkhasiat dalam mengatasi gangguan suasana hati dan

Judul Penelitian	Peneliti, tahun	Metode	Tujuan
Tidak Seimbang pada Pembuatan Model Komposisi Jamu			perilaku yang dihasilkan menggunakan regresi logistik melalui tahap SMOTE dengan model tanpa tahap SMOTE.
klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan <i>K-Nearest Neighbor</i>	Siringoringo (2018)	<i>Synthetic Minority Over-sampling Technique, k-Nearest Neighbor</i>	Mengatasi data tidak seimbang pada Dataset Credit Card Fraud serta metode k-Nearest Neighbor (kNN) untuk melaksanakan fungsi klasifikasi
<i>Research on unbalanced training samples based on SMOTE algorithm</i>	Li dan Hu (2019)	<i>SMOTE</i>	Meningkatkan kinerja klasifikasi pada sampel kelas minoritas pada data, yang mana akan menghasilkan kelas minoritas yang baru untuk menyeimbangkan data <i>training</i> .
<i>Directional Support Vector Machines</i>	Pernes, et al., (2019)	<i>directional statistics; supervised classification; support vector machines</i>	Dalam karya ini, peneliti merumuskan untuk pertama kalinya model non-probabilistik untuk klasifikasi data <i>directional</i> . Kami mengadopsi prinsip <i>max-margin</i> dan <i>the</i>

Judul Penelitian	Peneliti, tahun	Metode	Tujuan
			<i>hinge loss</i> , menghasilkan varian model <i>support vector machine</i> .
<i>Support Vector Machine</i>	Munoz, et al., (2019)	<i>Support Vector Machine</i>	Membuktikan di mana berbagai kernel dapat dikombinasikan dengan sejumlah algoritma pembelajaran yang berbeda, dan memiliki kinerja yang sangat baik pada data dengan dimensi tinggi.
Klasifikasi Nilai Mata Kuliah Berdasarkan Rekam Akademis Menggunakan <i>Support Vector Machine</i> : (Studi Kasus: Nilai Mata Kuliah Mahasiswa Statistika Universitas Islam Indonesia Angkatan 2015)	Amalia Dwi Nurfadzilah (2019)	<i>Statistika Deskriptif, Support Vector Machine, SMOTE</i>	Untuk mengetahui Statistika deskriptif dan Klasifikasi data nilai matakuliah mahasiswa Statistika UII angkatan 2015