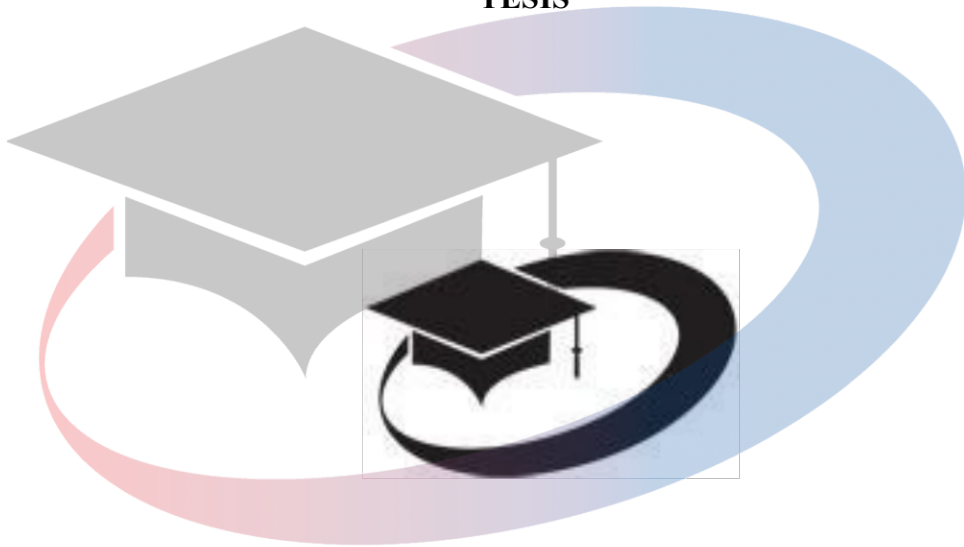


**METODE KNN PADA KLASIFIKASI RISIKO KREDIT
DENGAN SELEKSI FITUR BERBASIS BINARY PARTICLE
SWARM OPTIMIZATION**

TESIS

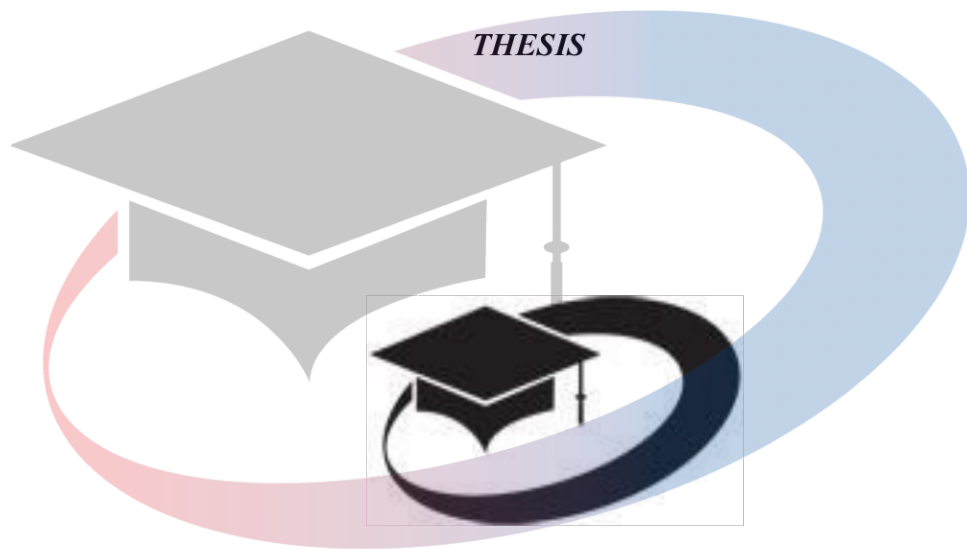


**UNIVERSITAS
MIKROSKIL**

Harmoko Lubis
NIM. 184211072

**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN
KOMPUTER
MIKROSKIL
MEDAN
2021**

***KNN METHOD FOR CREDIT RISK CLASSIFICATION USING
FEATURE SELECTION BASED BINARY PARTICLE SWARM
OPTIMIZATION***



**UNIVERSITAS
MIKROSKIL**

**Harmoko Lubis
NIM. 184211072**

**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN
KOMPUTER
MIKROSKIL
MEDAN
2021**

LEMBAR PENGESAHAN

**METODE KNN PADA KLASIFIKASI RISIKO KREDIT
DENGAN SELEKSI FITUR BERBASIS BINARY PARTICLE
SWARM OPTIMIZATION**

Diajukan untuk Melengkapi Persyaratan Guna

Mendapatkan Gelar Magister

Program Studi Magister Teknologi Informasi

Oleh :

HARMOKO LUBIS

NIM. 184211072

Disetujui Oleh:

Dosen Pembimbing I,

UNIVERSITAS
MIKROSKIL

Dr Pahala Sirait, S.T., M.Kom

Medan, 5 Februari 2021

Diketahui dan Disahkan Oleh:

Ketua Program Studi
Magister Teknologi Informasi,

Dr. Ronsen Purba, M.Sc

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “METODE KNN PADA KLASIFIKASI RISIKO KREDIT DENGAN SELEKSI FITUR BERBASIS BINARY PARTICLE SWARM OPTIMIZATION”.

Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi Magister Teknologi Informasi Strata Dua, STMIK Mikroskil Medan.

Penulis menyampaikan banyak terima kasih kepada:

1. Bapak Dr. Pahala Sirait, S.T., M.Kom., selaku Ketua STMIK Mikroskil Medan dan selaku Dosen Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
2. Bapak Dr. Ronsen Purba, M.Sc., selaku Ketua Program Studi Magister Teknologi Informasi yang telah membimbing penulis dalam proses penyelesaian tesis ini.
3. Bapak Arwin Halim S.Kom., M.Kom., selaku Dosen Pendamping yang telah membimbing penulis dalam proses penyelesaian tesis ini.
4. Bapak/Ibu Dosen STMIK Mikroskil yang telah membantu proses penulisan tesis ini.
5. Anggota keluarga, teman, saudara dan semua pihak yang terus memberikan dukungan penuh kepada penulis selama proses penulisan tesis ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan yang ada. Oleh sebab itu, kritik dan saran yang bersifat membangun akan sangat diterima. Akhir kata, semoga Tesis ini dapat bermanfaat bagi masyarakat. Terima kasih.

Medan, 05 Februari 2021

Penulis

Harmoko Lubis

184211072

LEMBAR PERNYATAAN

Saya yang membuat pernyataan ini adalah Mahasiswa Jurusan/Program Studi S-2 Magister Teknologi Informasi STMIK MIKROSKIL Medan dengan identitas mahasiswa sebagai berikut:

Nama : Harmoko Lubis
NIM : 184211072
Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul “METODE KNN PADA KLASIFIKASI RISIKO KREDIT DENGAN SELEKSI FITUR BERBASIS BINARY PARTICLE SWARM OPTIMIZATION”, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh STMIK Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada STMIK Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan Hak ini, STMIK Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, Februari 2021
Saya yang membuat pernyataan,



Harmoko Lubis

ABSTRAK

Saat ini, kinerja klasifikasi telah menjadi semakin penting untuk penilaian risiko kredit untuk pengendalian kerugian dan maksimalisasi pendapatan. Untuk itu, diperlukan metode klasifikasi yang dapat secara akurat dan efisien mengukur tingkat risiko kredit calon peminjam sebagai kunci proses persetujuan kredit. Penelitian ini berkontribusi pengembangan metode seleksi fitur dengan algoritma SI yang menggunakan representasi biner, yaitu seleksi fitur menggunakan algoritma PSO dengan representasi biner atau *Binary Particle Swarm Optimization* (BPSO) yang diterapkan pada klasifikasi risiko kredit, dengan evaluasi klasifikasinya menggunakan metode klasifikasi KNN. Penerapan pemilihan (seleksi) fitur dilakukan untuk menghilangkan fitur yang berlebihan, sehingga dapat mengurangi jumlah fitur, meningkatkan akurasi model, dan mengurangi waktu berjalan. Hasil pengujian menunjukkan bahwa akurasi terbaik KNN yaitu sebesar 76.40%, dapat ditingkatkan dengan fitur seleksi berbasis BPSO dengan perolehan akurasi yang lebih baik yaitu sebesar 88.70%, dengan peningkatan akurasi sebesar 13.35%. Pengujian ini menunjukkan bahwa teknik fitur seleksi berbasis BPSO berhasil meningkatkan akurasi klasifikasi KNN pada klasifikasi risiko kredit.

Kata Kunci: Klasifikasi Jantung Koroner, KNN, Fitur Seleksi, Binary PSO

UNIVERSITAS
MIKROSKIL

ABSTRACT

Classification performance has become increasingly important for credit risk assessment for loss control and revenue maximization. For this reason, a classification method is needed that can accurately and efficiently measure the credit risk level of a prospective borrower as the key to the credit approval process. This study contributed to the development of a feature selection method with the SI algorithm that uses binary representations, namely feature selection using the PSO algorithm with a binary representation or Binary Particle Swarm Optimization (BPSO) which is applied to credit risk classification, with evaluation of the classification using the KNN classification method. The application of feature selection is carried out to remove redundant features, thereby reducing the number of features, increasing model accuracy, and reducing run time. The test results show that the best accuracy of KNN is 76.40%, which can be improved by the BPSO-based selection feature with better accuracy of 88.70%, with an increase in accuracy of 13.35%. This test shows that the BPSO-based selection feature technique has succeeded in improving the accuracy of the KNN classification in the credit risk classification.

Keywords: Coronary Heart Classification, NPC, Features Selection, Binary PSO

UNIVERSITAS
MIKROSKIL

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	vii
BAB 1 KAJIAN LITERATUR	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	3
1.2.1 Identifikasi Masalah.....	3
1.2.1 Rumusan Masalah.....	4
1.3 Tujuan dan Manfaat Penelitian	4
1.4 Ruang Lingkup Penelitian.....	4
1.5 Metodologi Penelitian	5
1.6 Sistematika Penulisan	6
BAB 2 KAJIAN LITERATUR	7
2.1 Tinjauan Pustaka	7
2.1.1 <i>Data Mining</i>	7
2.1.2 <i>Unsupervised Learning</i>	9
2.1.2.1 <i>K-Means Clustering</i>	10
2.1.2.2 Hierarchical Clustering.....	12
2.1.2.3 Fuzzy C-Means.....	13
2.1.3 <i>Supervised Learning</i>	14
2.1.3.1 Decision Tree	15
2.1.3.2 <i>Artificial Neural Network</i>	16
2.1.3.3 <i>Support Vector Machine</i>	17
2.1.3.4 <i>Naive Bayes Classifier</i>	18
2.1.3.5 <i>k-Nearest Neighbor</i>	19

2.1.4	<i>Seleksi Fitur</i>	23
2.1.4.1	Seleksi Fitur menggunakan <i>Swarm Intelegent</i>	23
2.1.4.2	Seleksi Fitur Menggunakan PSO	24
2.1.4.3	Seleksi Fitur menggunakan Binary-PSO (BPSO).....	28
2.1.5	Evaluasi Klasifikasi	30
2.1.6	<i>k-Fold Cross Validation</i> (Metode Evaluasi Klasifikator).....	33
2.1.7	Kredit	34
2.1.8	Penelitian Terdahulu	40
2.2	Kerangka Pikir Pemecahan Masalah	41
BAB 3 METODOLOGI PENELITIAN		44
3.1	Analisis Masalah	44
3.2	Data yang Digunakan	46
3.3	Metode Penelitian.....	49
3.4	Alat Penelitian	56
3.5	Teknik Analisis	56
BAB 4 HASIL DAN PEMBAHASAN		58
4.1	Hasil	58
4.1.1	<i>Preprocessing Data</i>	58
4.1.1.1	Transformasi Data Kategorial	59
4.1.1.2	Normalisasi Data	61
4.1.2	Data Uji dan Data Latih	63
4.1.3	Penentuan nilai k untuk kNN	64
4.1.4	Klasifikasi KNN Tanpa Seleksi Fitur	64
4.1.5	Klasifikasi kNN dengan Seleksi Fitur Berbasis BPSO.....	68
4.1.5.1	Input Data dan Inisialisasi parameter	68
4.1.5.2	Membangun populasi awal.....	70
4.2	Pembahasan	75
BAB 5 KESIMPULAN DAN SARAN		94
5.1	Kesimpulan	94
5.2	Saran.....	94

DAFTAR GAMBAR

Gambar 3.1	Representasi standar kontiniu pada PSO untuk seleksi fitur.....	45
Gambar 3.2	Representasi standar biner pada PSO untuk seleksi fitur.....	46
Gambar 3.3	Flowchart Metode Penelitian	50
Gambar 3.4	Flowchart Seleksi Fitur dengan BPSO.....	54
Gambar 3.5	Flowchart Klasifikasi kNN dengan Fitur Seleksi berbasis BPSO ..	55
Gambar 4.1	Input Matriks Set Data Benchmark Credit German.....	62
Gambar 4.2	Hasil Normalisasi Data Benchmark Credit German	63
Gambar 4.3	Tampilan Hasil Klasifikasi BPSO-KNN.....	68
Gambar 4.4	Input Matriks Set Data	69
Gambar 4.5	Input Data dan Inisialisasi Parameter.....	70
Gambar 4.6	Tampilan Hasil Klasifikasi BPSO-NN.....	74
Gambar 4.7	Sampel Data Training 2 Fold	77
Gambar 4.8	Sampel Data Testing 2 Fold.....	77
Gambar 4.9	Sampel Data Training 5 Fold	80
Gambar 4.10	Sampel Data Testing 5 Fold.....	80
Gambar 4.11	Sampel Data Training 5 Fold	83
Gambar 4.12	Sampel Data Testing 5 Fold.....	83
Gambar 4.13	Sampel Data Training 10 Fold	85
Gambar 4.14	Sampel Data Testing 20 Fold.....	86

DAFTAR TABEL

Tabel 2.1	Confusion Matrix	21
Tabel 2.2	Penelitian Terdahulu	40
Tabel 3.1	Informasi Artribut	47
Tabel 4.1	Tampilan Set Data Benchmark	58
Tabel 4.2	Set Data Memiliki Data Bertipe Kategorial	59
Tabel 4.3	Transformasi Data Kategorial Menjadi Numerik	60
Tabel 4.4	Tampilan Set Data Setelah Transformasi.....	61
Tabel 4.5	Jarak Data Testing Data ₁ ke Setiap Data Training.....	66
Tabel 4.6	Data Latih Dengan Ranging Jarak Terkecil	66
Tabel 4.7	Dipilih Tetangga Data ₁ dengan asumsi Nilai k=3.....	67
Tabel 4.8	Representasi partikel ke i=1 PSO, membangkitkan partikel awal.....	71
Tabel 4.9	Susunan Biner Pada 50 Partikel Awal Yang Merepresentasikan Subset Fitur	71
Tabel 4.10	Perbandingan Akurasi KNN dan BPSOKNN dengan variasi pengujian nilai k pada validasi 2-fold	75
Tabel 4.11	Perbandingan Akurasi KNN dan BPSOKNN dengan variasi pengujian nilai k pada validasi 5-fold	78
Tabel 4.12	Perbandingan Akurasi KNN dan BPSOKNN dengan variasi pengujian nilai k pada validasi 10-fold	81
Tabel 4.13	Perbandingan Akurasi KNN dan BPSOKNN dengan variasi pengujian nilai k pada validasi 20-fold	84
Tabel 4.14	Perbandingan Akurasi Terbaik KNN dan BPSOKNN dengan variasi	86
Tabel 4.15	Hasil Perolehan Akurasi Pengujian Klasifikasi BPSOKNN dan KNN	87
Tabel 4.16	Perbandingan Presisi Pengujian Klasifikasi BPSOKNN dan KNN ..	88
Tabel 4.17	Perbandingan Recall Pengujian Klasifikasi BPSOKNN dan KNN ...	89
Tabel 4.18	Perbandingan F1 Score Klasifikasi BPSOKNN dan KNN	90
Tabel 4.19	Perbandingan Waktu Operasi Klasifikasi BPSOKNN dan KNN	92