RINGKASAN

Aditya Sofyan M. 2019. *Klasifikasi Calon Nasabah yang Berpotensi Melakukan Deposito Berjangka Menggunakan Algoritma C4.5 Berbasis Bagging*. Skripsi, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Malang. Pembimbing: (I) I Made Wirawan, S.T., S.ST., M.T. (II) Harits Ar Rosyid, S.T., M.T., Ph.D.

Kata Kunci : Klasifikasi*,* Deposito, C4.5, Bagging, SMOTE.

Deposito merupakan simpanan yang pengambilannya dilakukan dalam jangka waktu tertentu sesuai kesepakatan antara nasabah dengan pihak bank. Pada data distribusi simpanan bank umum yang dikeluarkan Lembaga Penjamin Simpanan periode sepanjang tahun 2018, deposito menjadi jenis simpanan dengan pendapatan terbesar dari total keseluruhan pendapatan bank. Hal tersebut membuktikan bahwa deposito sangat berpengaruh pada kehidupan sebuah bank karena mendapatkan nilai pemasukan paling besar. Maka dari itu, setiap bank perlu meningkatkan strategi pemasarannya untuk mencari calon nasabah yang berpotensi melakukan deposito. Pada penelitian ini dilakukan klasifikasi calon nasabah yang berpotensi melakukan deposito berjangka menggunakan algoritma C4.5 untuk meningkatkan strategi pemasaran *marketing* sebuah bank. Pada dataset yang digunakan terjadi ketidakseimbangan data*,* oleh karena itu diterapkan metode optimasi bagging untuk mengatasi ketidakseimbangan data dan meningkatkan nilai akurasi pada penelitian ini. Selain metode optimasi bagging, digunakan juga teknik SMOTE sebagai alternatif untuk mengatasi ketidakseimbangan data.

Proses klasifikasi menggunakan 2 model klasifikasi utama, yaitu menggunakan algoritma C4.5 dan algortima C4.5 berbasis bagging. Selain kedua model tersebut dilakukan juga 2 model klasifikasi sebagai alternatif, yaitu menggunakan algoritma C4.5 dengan SMOTE dan algoritma C4.5 berbasis bagging dengan SMOTE. Dan untuk evaluasi model klasifikasi diukur menggunakan nilai *precision, recall,* dan *f-measure* dari kelas *yes* serta nilai akurasi untuk mengukur kinerja klasifikasi keseluruhan.

Hasil klasifikasi menggunakan algoritma C4.5 mendapatkan nilai *precision* 57,71%, nilai *recall* 30,90%, nilai *f-measure* 40,25% dan nilai akurasi 89,43%. Setelah diterapkan metode optimasi bagging terhadap algoritma C4.5, didapatkan hasil yang tidak signifikan. Bagging dianggap tidak mampu mengatasi ketidakseimbangan data. Oleh karena itu dilakukan tambahan proses SMOTE sebagai solusi. Hasil yang didapat menunjukkan bahwa SMOTE meningkatkan nilai evaluasi kelas *yes* secara signifikan sehingga mendapatkan nilai yang tidak jauh beda dengan nilai akurasi.

SUMMARY

Aditya Sofyan M. 2019. *Classification of Potential Client with Term Deposit Using Bagging-Based C4.5*. Undergraduate Thesis, Departement of Electrical Engineering, Faculty of Engineering, State University of Malang. Advisor: (I) I Made Wirawan, S.T., S.ST., M.T. (II) Harits Ar Rosyid, S.T., M.T., Ph.D.

Keyword : Classification*,* Deposits, C4.5, Bagging, SMOTE.

Deposits are deposits that are taken in a certain period of time. In the data distribution of commercial bank deposits issued by the Deposit Insurance Corporation during the period of 2018, deposits are the type of savings with the largest income from the total income of the bank. This proves that deposits are very decisive in the life of a bank because it gets the highest income value. Therefore, each bank needs to improve its marketing strategy to find potential buyers who want to make a deposit. In this study, prospect analysis was conducted to discuss the provision of deposits using C4.5 to improve the marketing strategy of a bank. In the dataset used data imbalance occurs, therefore bagging optimization methods are applied to overcome imbalances class and increase the value of accuracy in this study. In addition to bagging optimization methods, the SMOTE technique is used as an alternative to overcome imbalance class.

The classification process uses 2 main classification models, namely using C4.5 algorithm and C4.5 algorithm based on bagging. In addition to the two models, 2 classification models were also used as alternatives, namely using the C4.5 algorithm with SMOTE and C4.5 based bagging algorithm with SMOTE. And for the evaluation of the classification model is measured using the value of precision, recall, and f-measure from the yes class and the value of accuracy to measure the overall classification performance.

The classification results using the C4.5 algorithm get a precision value of 57.71%, a recall value of 30.90%, a f-measure value of 40.25% and an accuracy value of 89.43%. After the bagging optimization method was applied to the C4.5 algorithm, the results were not significant. Bagging is considered unable to overcome imbalance class. Therefore an additional SMOTE process was carried out as a solution. The results obtained show that SMOTE significantly increases the evaluation value of the yes class so that it gets a value that is not much different from the value of accuracy.

# KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga proposal skripsi yang berjudul “Klasifikasi Calon Nasabah yang Berpotensi Melakukan Deposito Menggunakan Algoritma C4.5 Berbasis Bagging” dapat penulis selesaikan. Oleh karena itu pada kesempatan ini, penulis sampaikan ucapan terimakasih kepada:

1. Aji Prasetya Wibawa, S.T., M.M.T., Ph.D. selaku Ketua Jurusan Teknik Elektro Fakultas Teknik yang telah memberikan kebijakan dan sarana prasarana hingga terselesaikannya skripsi ini
2. Ilham Ari Elbaith Zaeni, S.T., M.T., Ph.D. selaku Kaprodi Teknik Informatika yang telah memberikan kebijakan dan sarana prasarana hingga terselesaikannya skripsi ini
3. I Made Wirawan, S.T., S.ST., M.T. selaku Dosen Pembimbing I yang telah memberikan arahan, motivasi, saran, waktu, dan banyak masukan berharga hingga terselesaikannya skripsi ini
4. Harits Ar Rosyid, S.T., M.T. Ph.D, selaku Dosen Pembimbing II yang telah memberikan arahan, motivasi, saran, waktu, dan banyak masukan berharga hingga terselesaikannya skripsi ini
5. Bapak Ibu Dosen Jurusan Elektro yang telah memberikan masukan dan ilmu yang bermanfaat dengan sabar dan tulus selama penulis menempuh perkuliahan
6. Kedua orang tua sebagai sumber motivasi utama yang telah mendukung, memberikan motivasi, dan doa yang tidak pernah terhenti kepada penulis sampai terselesaikannya skripsi ini.
7. Teman-teman Piggy yang telah memberikan motivasi, saran, waktu, dan arahan yang berharga hingga terselesaikannya skripsi ini.
8. Teman-teman TI-A yang telah memberikan motivasi dan waktu hingga terselesaikannya skripsi ini.

Semoga Tuhan Yang Maha Esa senantiasa melimpahkan rahmat dan perlindungan-Nya kepada semua pihak yang telah membantu hingga terselesaikannya skripsi ini. Penulis berharap tulisan ini bermanfaat bagi pembaca.

Malang, 26 Juli 2018

Penulis

**DAFTAR ISI**

Halaman

HALAMAN JUDUL ............................................................................................... i

LEMBAR PERSETUJUAN ................................................................................... ii

LEMBAR PENGESAHAN ................................................................................... iii

PERNYATAAN KEASLIAN TULISAN ............................................................. iv

ABSTRAK v

SUMMARY .......................................................................................................... vi

KATA PENGANTAR vii

DAFTAR ISI viii

DAFTAR TABEL xi

DAFTAR GAMBAR xii

BAB I PENDAHULUAN

1.1 Latar Belakang 1

1.2 Rumusan Masalah 3

1.3 Tujuan Penelitian 3

1.4 Batasan Masalah 3

1.5 Manfaat Penelitian 4

BAB II KAJIAN PUSTAKA

2.1 Bank dan Perbankan 5

2.2 Sumber-Sumber Dana Bank 5

2.3 Deposito 6

2.4 Klasifikasi 7

2.5 *Decision Tree* (Algoritma C4.5) 7

2.6 *Synthetic Minority Over-sampling Technique* (SMOTE) 11

2.7 *Bootstrap Aggregating* (Bagging) ............................................... 12

2.8 *K-Fold Cross Validation* ............................................................. 13

2.9 *Confusion Matrix* ........................................................................ 14

2.10 Penelitian Relevan ...................................................................... 15

2.11 Kerangka Berpikir ...................................................................... 16

BAB III METODE PENELITIAN

3.1 Desain Penelitian 20

3.2 Studi Literatur ............................................................................ 20

3.3 Pengumpulan dan Analisis Data ................................................ 21

3.4 Pengolahan Data Awal ............................................................... 22

3.5 Proses Klasifikasi ....................................................................... 22

3.5.1 Model Klasifikasi Algoritma C4.5 ................................... 23

3.5.2 Model Klasifikasi Algoritma C4.5 Berbasis Bagging ...... 24

3.5.3 Model Klasifikasi Algoritma C4.5 dengan SMOTE ........ 26

3.5.4 Model Klasifikasi Algoritma C4.5 Berbasis Bagging

dan SMOTE ..................................................................... 28

3.6 Validasi dan Evaluasi ................................................................. 30

3.7 Contoh Perhitungan Manual ...................................................... 31

3.7.1 *Discretization* ................................................................... 32

3.7.2 SMOTE ............................................................................ 33

3.7.3 *Bootstrap* .......................................................................... 34

3.7.4 Klasifikasi Algoritma C4.5 .............................................. 36

3.7.5 *Aggregating* ...................................................................... 41

3.7.6 Evaluasi Hasil ................................................................... 43

BAB IV HASIL DAN PEMBAHASAN

4.1 Preprocessing Data ..................................................................... 46

4.1.1 *Discretization* .................................................................... 46

4.1.2 SMOTE ............................................................................. 46

4.2 Pengujian Algoritma ................................................................... 47

4.2.1 Pengujian Algoritma C4.5 ................................................ 47

4.2.2 Pengujian Algoritma C4.5 Berbasis Bagging ................... 49

4.2.3 Pengujian Algoritma C4.5 dengan SMOTE ..................... 50

4.2.4 Pengujian Algoritma C4.5 Berbasis Bagging

dengan SMOTE ................................................................ 52

4.3 Perbandingan Hasil Algoritma C4.5 dan Algoritma C4.5

Berbasis Bagging ........................................................................ 56

4.4 Pengaruh SMOTE Terhadap Hasil Klasifikasi Algoritma C4.5

dan Algoritma C4.5 Berbasis Bagging ........................................ 60

BAB V PENUTUP

5.1 Kesimpulan .................................................................................... 59

5.2 Saran .............................................................................................. 59

DAFTAR PUSTAKA 60

RIWAYAT HIDUP 63

LAMPIRAN .......................................................................................................... 64

DAFTAR GAMBAR

Gambar Halaman

2.1 Jenis Teknik Klasifikasi 7

2.2 Arsitektur *Decision Tree*  8

2.3 Flowchart Algoritma C4.5 9

2.4 Alur Teknik SMOTE 11

2.5 Alur Metode Bagging 12

2.6 Kerangka Penelitian Pertama 17

2.7 Kerangka Penelitian Kedua 18

2.8 Kerangka Penelitian Ketiga 18

2.9 Kerangka Penelitian Keempat 19

3.1 Tahapan Penelitian 20

3.2 Flowchart Algoritma C4.5 23

3.3 Flowchart Algoritma C4.5 dengan SMOTE 25

3.4 Flowchart Algoritma C4.5 Berbasis Bagging 27

3.5 Flowchart Algoritma C4.5 Berbasis Bagging dengan SMOTE 29

3.6 Atribut *Duration* sebagai Akar Simpul 38

3.7 Cabang Atribut Duration > 1114 40

3.8 Pohon Hasil Klasifikasi C4.5 41

3.9 Pohon Model 1-6 Hasil Bagging 41

3.10 Pohon Model 7-10 Hasil Bagging 42

4.1 Hasil Pengujian Algoritma C4.5 47

4.2 Hasil Pengujian Algoritma C4.5 Berbasis Bagging49

4.3 Perbandingan Nilai Klasifikasi Algoritma C4.5 dan

Algoritma C4.5 Berbasis Bagging 49

4.4 Hasil Pengujian Algoritma C4.5 dengan SMOTE 52

4.5 Hasil Pengujian Algoritma C4.5 Berbasis Bagging dengan SMOTE54

4.9 Perbandingan Nilai *Precision* dengan bagging dan SMOTE 55

4.10 Perbandingan Nilai *Recall* dengan bagging dan SMOTE 56

4.11 Perbandingan Nilai *F-Measure* dengan bagging dan SMOTE 56

4.12 Perbandingan Nilai *Akurasi* dengan bagging dan SMOTE 57

DAFTAR TABEL

Tabel Halaman

2.1 Skema *10-Fold Cross Validation* 13

2.2 *Confusion Matrix* 14

2.3 Rangkuman Penelitian Relevan 15

3.1 Keterangan Data21

3.2 Data Bank *Marketing* (Sampel 1-5) 31

3.3 Data Bank *Marketing* (Sampel 6-10) 31

3.4 Atribut *Age* 32

3.5 Hasil Contoh Diskritisasi Data 33

3.6 Data Bank *Marketing* Setelah Proses SMOTE 33

3.7 Tabel Hasil *Bootstrap* 34

3.8 Hasil Perhitungan Semua Atribut 36

3.9 Sampel Data Atribut *Duration <*1114 38

3.10 Sampel Data Atribut *Duration >*1114 39

3.11 Hasil Proses *Agregating* 42

3.12 Hasil Evaluasi 43

3.13 *Confusion Matrix* 43

4.1 Hasil Diskretisasi Data 45

4.2 Nilai *Confusion Matrix* Algoritma C4.5 46

4.3 Nilai *Confusion Matrix* Algoritma C4.5 Berbasis Bagging47

4.4 Nilai *Confusion Matrix* Algoritma C4.5 dengan SMOTE 51

4.5 Nilai *Confusion Matrix* Algoritma C4.5 Berbasis Bagging

dan SMOTE 53