

**ANALISIS DAN KOMPARASI ALGORITMA MACHINE LEARNING
UNTUK ESTIMASI WAKTU DAN BIAYA PENGEMBANGAN
PERANGKAT LUNAK**

TESIS

**Diajukan Sebagai Salah Satu Syarat Untuk Menyelesaikan
Program Strata Dua (S2) Magister Komputer**



OLEH:
MISNO
39111221084

**PROGRAM STUDI TEKNIK INFORMATIKA
PROGRAM PASCA SARJANA (S2) MAGISTER KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER ERESHA
JAKARTA
2014**

PERSETUJUAN TESIS

Nama : MISNO
NPM : 390111221084
Konsentrasi : SOFTWARE ENGINEERING
Judul tesis : Analisis dan Komparasi Algoritma *Machine Learning* untuk
Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak

Telah disetujui untuk disidangkan pada Sidang Tesis pada Program Pasca Sarjana (S2)
Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen
Informatika dan Komputer Eresha.

Jakarta, 25 September 2013

Pembimbing Utama

Pembimbing Pendamping

Romi Satria Wahono, M.Eng, Ph.D.

Bobby Reza, S.Kom, MM.

Mengetahui

Ketua

Direktur

Program Pasca Sarjana

Ir. Damsiruddin Siregar, MMT

Dr. Rufman Iman Akbar E.,MM., M.Kom

PENGESAHAN TESIS

Nama : MISNO
NPM : 390111221084
Konsentrasi : SOFTWARE ENGINEERING
Judul tesis : Analisis dan Komparasi Algoritma *Machine Learning* untuk
Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak

Telah disidangkan dan dinyatakan Lulus Sidang Tesis pada Program Pasca Sarjana (S2)
Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen
Informatika dan Komputer Eresha pada tanggal 09 Pebruari 2013.

Nama Penguji

Tanda Tangan

.....
(Ketua)

.....

.....
(Sekretaris)

.....

.....
(Anggota)

.....

Mengetahui:

Direktur

Program Pasca Sarjana

Dr. Rufman Iman Akbar E.,MM., M.Kom

PERNYATAAN KEASLIAN TESIS

Nama : Misno
NPM : 390111221084
Konsentrasi : Software Engineering
Judul Tesis : Analisis dan Komparasi Algoritma *Machine Learning* untuk
Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak

Dengan ini menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang sepengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Jakarta, 20 September 2014

Misno

Misno, 390111221084

Analisis dan Komparasi Algoritma Machine Learning Untuk Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak; dibawah pembimbing Romi Satria Wahono, M.Eng, Ph.D. dan Bobby Reza, S.Kom, MM.

103 + xiv hal / 20 tabel / 11 gambar / 3 lampiran / 28 pustaka (1998 – 2014)

ABSTRAK

Pengembangan perangkat lunak telah menjadi kegiatan penting bagi banyak organisasi modern. Bahkan kualitas, biaya, dan ketepatan waktu perangkat lunak yang dikembangkan sering penentu penting dari kesuksesan organisasi. Ada implikasi keuangan dan strategis mendasar di antara keduanya untuk proyek-proyek pembangunan dalam hal aktivitas penjadwalan dan estimasi biaya. Estimasi biaya Software adalah salah satu tugas yang paling penting dalam mengelola proyek perangkat lunak. Biaya pengembangan cenderung meningkat dengan kompleksitas proyek dan perkiraan biaya maka akurat sangat diinginkan selama pengembangan tahap awal. Estimasi biaya dan waktu merupakan faktor penting untuk suksesnya proyek-proyek pembangunan perangkat lunak. Banyak metode yang telah diusulkan untuk estimasi biaya dan waktu pengembangan perangkat lunak. SLIM, COCOMO, *Function Point*, *Use Case Point* merupakan metode konvensional yang sering digunakan untuk estimasi. Selain metode konvensional diatas, juga banyak diusulkan metode machine learning. Metode machine learning banyak diusulkan dan dipercaya memberikan hasil lebih akurat jika dibandingkan metode konvensional. Algoritma *machine learning* yang banyak digunakan diantaranya *Neural Network*, *Support Vector Maching* dan *Linear Regression*. Dari pengujian dengan mengukur kinerja ketiga metode tersebut untuk beberapa dataset, disimpulkan bahwa *Support Vector Machine* memiliki akurasi paling tinggi atau prediksi error paling sedikit, diukur dengan menggunakan matrix RMSE, diikuti oleh *Neural Network* dan *Linear Regression*.

Kata kunci: Estimasi biaya dan waktu proyek, *machine learning* untuk estimasi, *Support Vector Machine*, *Neural Network*, *Linear Regression*

Misno, 390111221084

Analysis and Comparison of Machine Learning Algorithm for Software Effort Estimation;
under the guidance of Romi Satria Wahono, M.Eng, Ph.D. and Bobby Reza, S.Kom, MM.
103 + xiv pages / 20 tables / 11 pictures / 3 attachments / 28 citations (1998 – 2012)

ABSTRACT

Software development has become an important activity for many modern organizations. In fact the quality, cost, and timeliness of developed software are often crucial determinants of an organization's success. There are significant financial and strategic implications for development projects in terms of activity scheduling and cost estimation. Software cost estimation is one of the most critical tasks in managing software projects. Development costs tend to increase with project complexity and hence accurate cost estimates are highly desired during the early stages development. Accurate estimate in cost and time is an important factor for the success of software development projects. Many methods have been proposed to estimate the cost and time of software development. SLIM, COCOMO, Function Point, Use Case Point are some of conventional method that is often used for estimation. In addition to the above conventional methods, many machine learning method also have been proposed. Many machine learning method is proposed and is believed to give more accurate results than conventional methods. Many machine learning method is proposed and are believed to give more accurate results than conventional methods. Machine learning algorithms are widely used for example Neural Network, Support Vector Maching dan Linear Regression. By experiment, measure the performance of the three methods for several datasets, concluded that Support Vector Machine has the highest accuracy of prediction error or at least, measured using RMSE matrix, followed by a Neural Network and Linear Regression.

Keywords: Project cost and time estimation, machine learning for estimation, Support Vector Machine, Neural Network, Linear Regression

KATA PENGANTAR

Dengan mengucap puji syukur kehadiran Allah Swt. yang telah melimpahkan rahmat dan hidayah-Nya kepada penulis, sehingga penulis dapat menyelesaikan tesis yang berjudul ” Analisis dan Komparasi Algoritma *Machine Learning* Untuk Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak”. Ucapan terima kasih juga saya sampaikan kepada orang tua, istri, anak serta kakak dan adik-adik saya yang telah memberikan do’a restu dan dorongan moral.

Tesis ini digunakan untuk melengkapi salah satu persyaratan yang diajukan dalam rangka menempuh ujian akhir untuk memperoleh gelar Magister Komputer (M. Kom.) pada Program Pasca Sarjana (S2) Program Studi Teknik Informatika di Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.

Penulis menyadari bahwa penyusunan tesis ini tidak akan berhasil tanpa adanya dukungan dan bantuan dari berbagai pihak. Maka dalam kesempatan ini penulis memberikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Ir. Damsirudin Siregar, M.MT, selaku Ketua STMIK Eresha Jakarta, yang telah memberikan kesempatan kepada penulis untuk belajar dan menuntut ilmu pada almamater yang beliau pimpin.
2. Bapak Dr. Rufman Imam Akbar Effendy, SE., MM., M. Kom, selaku Direktur Program Pasca Sarjana Magister Komputer STMIK Eresha Jakarta, sekaligus sebagai dosen pembimbing yang telah banyak memberikan masukan dan koreksi tesis ini.
3. Bapak Didik Setyadi, M. Kom, sebagai Puket II STMIK Eresha Jakarta, yang telah banyak memberikan motivasi kepada penulis untuk menyelesaikan tesis ini.
4. Bapak Bobby Reza, S. Kom., MM. selaku Puket III STMIK Ersha Jakarta, yang telah banyak memberikan motivasi dan dorongan kepada penulis untuk menyelesaikan tesisi ini.
5. Bapak Romi Satria Wahono, M. Eng., Ph.D, selaku dosen pembimbing utama, yang telah banyak memberikan arahan dan bimbingannya sehingga penulis dapat menyelesaikan tesis ini.
6. Seluruh dosen dan staf STMIK Eresha Jakarta, yang telah memberikan dukungan moril dan bantuan dalam proses pembelajaran.

7. Rekan-rekan angkatan 39 yang telah banyak memberikan bantuan dan dukungan moral sehingga proses penyusunan tesis ini selesai.
8. Orang tua, isteri, anak-anak saya semua yang telah memberikan dukungan dalam terselesaikannya penulisan tesis ini.

Akhir kata penulis mohon maaf atas kekeliruan dan kesalahan yang terdapat dalam tesis ini dan berharap semoga tesis ini dapat bermanfaat bagi khasanah pengetahuan teknologi informasi di Indonesia.

Penulis

DAFTAR ISI

| | |
|--|-------|
| | Hal |
| PERSETUJUAN TESIS | ii |
| PENGESAHAN TESIS | iii |
| PERNYATAAN KEASLIAN TESIS | iv |
| ABSTRAK..... | v |
| ABSTRACT..... | vi |
| KATA PENGANTAR | vii |
| DAFTAR ISI..... | ix |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL..... | xiii |
| DAFTAR LAMPIRAN..... | xiv |
| Bab 1 Pendahuluan | 1 |
| 1.1 Latar Belakang Masalah..... | 1 |
| 1.2 Identifikasi Masalah | 4 |
| 1.3 Rumusan Masalah | 4 |
| 1.4 Ruang Lingkup Masalah | 4 |
| 1.5 Tujuan Penelitian..... | 5 |
| 1.6 Manfaat Penelitian..... | 5 |
| 1.7 Sistematika Penulisan..... | 5 |
| Bab 2 Landasan Teori..... | 7 |
| 2.1 Tinjauan Pustaka | 7 |
| 2.1.1 Model Penelitian Nassif, Danny Ho, & Luiz Fernando Capretz | 7 |
| 2.1.2 Model Penelitian de Barcelos Tronto | 8 |
| 2.1.3 Model Penelitian Oliveira | 10 |
| 2.1.4 Rangkuman Penelitian Terkait..... | 11 |
| 2.2 Landasan Teori..... | 12 |

| | | |
|-------|--|----|
| 2.2.1 | Pengembangan Perangkat Lunak | 13 |
| 2.2.2 | Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak | 14 |
| 2.3 | Algoritma Machine Learning | 20 |
| 2.3.1 | Algoritma Support Vector Machine (SVM) | 23 |
| 2.3.2 | Algoritma Neural Networks (NN) | 28 |
| 2.3.3 | Algoritma Genetika..... | 33 |
| 2.3.4 | Particle Swarm Optimization (PSO) | 38 |
| 2.4 | Metode Komparasi dan Evaluasi Algoritma Machine Learning untuk Estimasi Biaya dan Waktu Pengembangan Perangkat Lunak | 40 |
| 2.5 | Kerangka Pemikiran Penelitian | 42 |
| Bab 3 | Metode Penelitian..... | 44 |
| 3.1 | Perancangan Penelitian | 44 |
| 3.2 | Teknik Analisis | 45 |
| 3.2.1 | Pengumpulan Data | 45 |
| 3.2.2 | Pengolahan Awal Data..... | 48 |
| 3.2.3 | Model yang Diusulkan..... | 55 |
| 3.2.4 | Eksperimen dan Pengujian Model | 57 |
| 3.2.5 | Evaluasi dan Validasi Hasil | 57 |
| Bab 4 | Hasil dan Pembahasan..... | 58 |
| 4.1 | Hasil | 58 |
| 4.1.1 | Hasil Pengitungan Manual | 58 |
| 4.1.2 | Hasil Pengujian dan Komparasi Algoritma Estimasi..... | 66 |
| 4.2 | Pembahasan..... | 66 |
| 4.2.1 | Parametric test..... | 67 |
| 4.2.2 | Non-Parametric test | 67 |
| 4.3 | Analisis Hasil Komparasi..... | 67 |
| 4.4 | Implikasi Penelitian..... | 68 |

| | | |
|----------------|---------------------------|----|
| Bab 5 | Kesimpulan dan Saran..... | 69 |
| 5.1 | Kesimpulan..... | 69 |
| 5.2 | Saran..... | 69 |
| Daftar PUSTAKA | | 70 |

DAFTAR GAMBAR

| | Hal |
|--|-----|
| Gambar 2.1 Model yang diusulkan oleh Ali Bou Nassif | 8 |
| Gambar 2.2 Model yang diusulkan oleh Barcelos Tronto | 9 |
| Gambar 2.3 Model yang diusulkan oleh Oliveira | 11 |
| Gambar 2.4 Trilpe Constrain Project | 14 |
| Gambar 2.5 Flowchart algoritma genetika | 36 |
| Gambar 2.6 Kerangka Pemikiran | 43 |
| Gambar 3.1 Rancangan model yang diusulkan | 56 |

DAFTAR TABEL

| | Hal |
|---|-----|
| Tabel 2-1 Perbandingan penelitian serupa | 12 |
| Tabel 2-2 Fungsi Kernel | 26 |
| Tabel 3-1 Keterangan Atribut COCOMO..... | 45 |
| Tabel 3-2 Dataset COCOMO 81..... | 47 |
| Tabel 3-3 Dataset COCOMONASA 2..... | 47 |
| Tabel 3-4 Dataset COCOMONASA 1..... | 47 |
| Tabel 3-5 Daftar Preprocessing Data | 50 |
| Tabel 3-6 Nilai dari <i>Multiplier</i> COCOMO | 50 |
| Tabel 3-7 Dataset COCOMO81 awal | 52 |
| Tabel 3-8 Dataset NASA Awal..... | 52 |
| Tabel 3-9 Dataset COCOMO81 setelah transformasi numerik | 53 |
| Tabel 3-10 Dataset NASA93 Awal..... | 53 |
| Tabel 3-11 Hasil Pengolahan awal dataset NASA93 | 54 |
| Tabel 3-12 Hasil Pengolahan awal dataset COCOMO81 | 54 |
| Tabel 3-13 Hasil Pengolahan awal dataset NASA | 54 |
| Tabel 3-14 Spesifikasi komputer yang digunakan..... | 57 |
| Tabel 4-1 Sampel dataset COCOMO81 | 59 |
| Tabel 4-3 Rangkuman dari nilai RMSE percobaan LR, NN dan SVM..... | 67 |

DAFTAR LAMPIRAN

| | |
|--|------------|
| Lampiran 1 Dataset Cocomo81 | 76 |
| Lampiran 2 Dataset Cocomo81 Logaritma Natural..... | 78 |
| Lampiran 3 Dataset Cocomonasa 1 | 81 |
| Lampiran 4 Dataset NASA Tansformasi Numerik..... | 84 |
| Lampiran 5 Dataset NASA Logaritma Natural | 86 |
| Lampiran 6 Dataset NASA93 | 89 |
| Lampiran 7 Dataset NASA93 Select Attribut | 101 |
| Lampiran 8 Dataset NASA93 Transformasi Numerik..... | 103 |
| Lampiran 9 Dataset NASA93 Logaritma Natural | 106 |

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Pengembangan perangkat lunak adalah kegiatan profesional di mana perangkat lunak dikembangkan untuk tujuan bisnis yang spesifik, untuk dimasukkan dalam perangkat lain, atau sebagai produk perangkat lunak seperti sistem informasi, sistem CAD dan sebagainya (Sommerville, 2010). Ada tiga parameter suksesnya proyek pengembangan perangkat lunak (Schwalbe, 2013):

1. Proyek memenuhi target *scope*, waktu dan biaya
2. Hasil proyek pengembangan dapat memuaskan pelanggan/sponsor
3. Hasil proyek pengembangan memenuhi tujuan utama proyek pengembangan perangkat lunak.

Kesuksesan proyek pengembangan dipengaruhi oleh banyak hal, diantaranya dukungan esekutif, keterlibatan pengguna dalam proyek, pengalaman manager proyek, tujuan bisnis yang jelas, infrastruktur perangkat lunak yang str dan penggunaan metodologi pengembangan yang formal. Faktor lainnya merupakan faktor yang berkaitan dengan *timing* dan *scope* proyek, diantaranya *scope* yang minimal dan estimasi yang handal (Schwalbe, 2013). Estimasi sumber daya, biaya, dan jadwal untuk pengembangan perangkat lunak membutuhkan pengalaman, akses ke informasi sejarah proyek yang baik, dan keberanian untuk melakukan prediksi kuantitatif sedangkan informasi kualitatif adalah semua fakta yang ada. Estimasi membawa risiko yang melekat dan risiko ini dapat menyebabkan ketidakpastian dalam proyek.

Estimasi waktu dan biaya pengembangan perangkat lunak adalah proses melakukan estimasi biaya dan waktu yang diperlukan untuk pengembangan perangkat lunak (Wen, Li, Lin, Hu, & Huang, 2012). Perusahaan-perusahaan pengembang perangkat lunak banyak yang dihadapkan dengan masalah estimasi biaya dan waktu yang diperlukan untuk proyek pengembangan perangkat lunak (The Standish Group, 2009). Estimasi jadwal proyek sulit dilakukan karena

banyak hal, misalnya perangkat lunak yang dikembangkan kemungkinan harus berjalan di lingkungan sistem yang sama sekali berbeda, menggunakan pengembangan dengan teknologi baru atau kompetensi orang-orang yang terlibat dalam proyek yang mungkin tidak diketahui. Ada begitu banyak ketidakpastian sehingga sulit untuk memperkirakan biaya dan waktu pengembangan secara akurat selama awal tahapan proyek (Sommerville, 2010). Biaya pengembangan perangkat lunak cenderung meningkat seiring dengan kompleksitas proyek dan oleh karena itu, estimasi biaya yang akurat sangat diperlukan dalam tahap awal proyek pengembangan perangkat lunak (Xu & Khoshgoftaar, 2004). Estimasi yang akurat merupakan salah satu faktor yang mempengaruhi kesuksesan proyek pengembangan perangkat lunak (Schwalbe, 2013), dan merupakan suatu hal yang vital untuk mencegah kegagalan atau macetnya pengembangan perangkat lunak (Menzies & Hihn, 2006). Oleh karena itu, diperlukan suatu metode dalam melakukan estimasi biaya dan waktu dalam proyek pengembangan perangkat lunak supaya hasilnya dapat diandalkan.

Sejak tahun 1980-an, banyak metode estimasi telah diusulkan untuk estimasi biaya dan waktu pengembangan perangkat lunak. Beberapa metode yang telah diusulkan dalam estimasi biaya dan waktu pengembangan perangkat lunak diantaranya SLIM (Putnam, 1978), COCOMO (Boehm, 1981), Expert Judgement (Jorgensen & Kirkeboen, 2000), Function Point Analysis (Albrecht & Gaffney, 1983) dan Use Case Point (Karner, 1993) (Anda, Angelvik, & Ribu, 2002). COCOMO (Constructive Cost Model), diusulkan oleh Barry Boehm (Boehm, 1981), adalah metode yang paling populer dalam estimasi biaya dan waktu pengembangan perangkat lunak. Metode ini menggunakan beberapa persamaan dan parameter, yang telah diperoleh dari pengalaman estimasi biaya dan waktu proyek sebelumnya. COCOMO II merupakan versi terbaru dari COCOMO yang melakukan estimasi biaya dan waktu berdasarkan dari *person-month*(PM) dalam proyek perangkat lunak.(Dejaeger, Verbeke, Martens, & Baesens, 2012)

Metode-metode estimasi biaya dan waktu pengembangan perangkat lunak yang disebutkan sebelumnya merupakan metode konvensional yang memiliki tingkat keakuratan yang relatif rendah (Wen et al., 2012). Selain metode-metode konvensional di atas, belakangan ini dikembangkan estimasi dengan algoritma machine learning. Beberapa algoritma machine learning yang telah digunakan

untuk estimasi biaya dan waktu pengembangan perangkat lunak diantaranya adalah Case-Based Reasoning (Brady, 2010), Neural Networks (Araújo, Oliveira, Soares, & Meira, 2012) (Pendharkar, 2010), Bayesian Statistics, Linear Regressions (Elish, 2009), Genetic Programming (Afzal & Torkar, 2011).

Beberapa peneliti telah melakukan komparasi menggunakan beberapa algoritma pada dataset *Software Effort*. Penelitian Nassif et al (Nassif, Ho, & Capretz, 2013) membandingkan algoritma LR dengan ANN dengan menggunakan dataset ISBG, CompuTop dan Western University - Canada, didapatkan hasil terbaik adalah ANN. Penelitian De Barcelos et al (de Barcelos Tronto, da Silva, & Sant'Anna, 2008) membandingkan algoritma ANN dan LR dengan menggunakan dataset COCOMO81 dan didapatkan hasil terbaik adalah ANN. Penelitian Oliveira (Oliveira, 2006) membandingkan algoritma RBFN dengan SVR menggunakan dataset NASA, didapatkan hasil terbaik adalah SVR.

Dalam penelitian tersebut, juga digunakan banyak dataset diantaranya cocomo81 (Brady, 2010), nasa (Brady, 2010), ISBG (Brady, 2010), Desharnais dan Maxwell (Brady, 2010), IBM DP (Vinay Kumar, Ravi, Carr, & Raj Kiran, 2008) (Heiat, 2002), Korean IT Vendor (Park & Baek, 2008). Dataset software effort banyak memiliki atribut yang beragam. Misalnya dataset Albrecht memiliki 8 atribut, dataset cocomo81 memiliki 16 atribut dan dataset ISBG memiliki 91 atribut. Kinerja yang lebih baik dapat dicapai dengan menghilangkan beberapa atribut, yang dengan ini dapat menghilangkan *noise* dan atribut yang tidak relevan (Zhao, Fu, Ji, Tang, & Zhou, 2011). *Feature selection* dapat digunakan untuk menentukan *subset of features* yang optimal, dimana hal ini dapat meningkatkan akurasi estimasi (Chen, Menzies, Port, & Boehm, 2005).

Beberapa metode *feature selection* telah digunakan dalam estimasi biaya dan waktu pengembangan perangkat lunak, diantaranya principle component analysis (PCA) (Wen, Li, & Tang, 2009), Backward Elimination (Chen et al., 2005), Backward Elimination (Chen et al., 2005), Particle Swarm Optimization (Liu et al., 2011) (Khatibi Bardsiri, Jawawi, Hashim, & Khatibi, 2012) (Huang & Dun, 2008), dan Genetic Algorithm (Braga, Oliveira, & Meira, 2008). Dari beberapa model *feature selection* yang digunakan, belum diketahui metode yang paling baik dalam estimasi biaya dan waktu pengembangan perangkat lunak.

Dari semua hasil penelitian yang sudah dilakukan belum ada satupun model estimasi yang terbukti paling akurat dengan menggunakan semua dataset yang ada. Oleh karena itu, pada penelitian ini akan dilakukan komparasi beberapa algoritma machine learning (LR, NN dan SVM) dan algoritma *feature selection* yang terbaik pada semua dataset *software effort*.

1.2 Identifikasi Masalah

Berdasarkan uraian dari latar belakang maka didapati permasalahan sebagai berikut:

Dataset software effort jumlahnya banyak dan memiliki atribut yang beragam, jika semua atribut pada dataset tersebut digunakan, maka akan menurunkan *accuracy* hasil estimasi. Disamping itu, banyak metode *machine learning* dan *dataset* yang telah digunakan untuk estimasi biaya dan waktu pengembangan perangkat lunak, akan tetapi belum ada komparasi yang komprehensif untuk semua *dataset* yang ada untuk menentukan algoritma mana yang paling baik, untuk melakukan estimasi biaya dan waktu pengembangan perangkat lunak.

1.3 Rumusan Masalah

Dari identifikasi masalah diatas, rumusan masalah penelitian ini adalah:

Algoritma machine learning apa yang paling baik untuk estimasi biaya dan waktu pengembangan perangkat lunak.

1.4 Ruang Lingkup Masalah

Ruang lingkup masalah dalam penelitian ini hanya membahas tentang perbandingan metode *machine learning* beserta dataset yang digunakan dalam estimasi biaya dan waktu pengembangan perangkat lunak.

1.5 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Untuk mengidentifikasi algoritma machine learning yang paling akurat untuk estimasi biaya dan waktu pengembangan perangkat lunak.
2. Untuk mengidentifikasi algoritma *parameter selection* yang paling baik untuk estimasi biaya dan waktu pengembangan perangkat lunak.

1.6 Manfaat Penelitian

Diharapkan, hasil penelitian ini dapat dimanfaatkan untuk pengambilan keputusan dalam melakukan estimasi pengembangan perangkat lunak dan dapat menjadi rujukan untuk penelitian tentang estimasi biaya dan waktu pengembangan perangkat lunak selanjutnya.

1.7 Sistematika Penulisan

Sistematika penulisan untuk penelitian ini terbagi dalam 5 bab seperti yang tertera di bawah ini.

Bab I Pendahuluan

Pada bab ini akan membahas tentang latar belakang, permasalahan penelitian, tujuan dan manfaat penelitian, serta sistematika penulisan dari penelitian ini.

Bab II Landasan Teori

Pada bab ini berisi tentang teori yang melsi penelitian yaitu mengenai perangkat lunak, estimasi perangkat lunak, dan metode estimasi dengan

algoritma machine learning yang sering digunakan dalam estimasi biaya dan waktu pengembangan perangkat lunak.

Bab III Metodologi Penelitian

Pada bab ini membahas tentang metode pengumpulan data dan metode penelitian. Penelitian di sini adalah membuat model dengan algoritma machine learning untuk estimasi biaya dan waktu pengembangan perangkat lunak.

Bab IV Analisa Dan Hasil Pembahasan

Pada bab ini menjelaskan tentang hasil dan pembahasan penelitian dari penelitian yang dilakukan. Hasil merupakan suatu penjelasan tentang data kuantitatif yang dikumpulkan dari pengolahan dataset yang digunakan. Pembahasan merupakan suatu penjelasan tentang pengolahan data dan interpretasinya, baik dalam bentuk diskriptif ataupun penarikan inferensinya. Pada bab ini dilakukan pengujian model yang dihasilkan dari bab sebelumnya. Pengujian dilakukan dengan mengukur kinerja tiap algoritma menggunakan beberapa metode pengujian kemudian hasil pengukurannya dikomparasi untuk melihat akurasi dari algoritma-algoritma tersebut, kemudian dilakukan validasi dalam hasil estimasi yang didapat.

Bab V Penutup

Pada bab ini berisi ringkasan temuan, rangkuman kesimpulan dan saran. Kesimpulan merupakan pernyataan secara general atau spesifik yang berisi hal-hal penting dan menjadi temuan penelitian yang bersumber pada hasil dan pembahasan. Saran merupakan pernyataan atau rekomendasi peneliti yang berisi hal-hal penting sebagaimana yang telah disampaikan yaitu implikasi penelitian.