

Prediksi Kemampuan *Sprint Delivery* dalam *Iterative-based Software Development*

Clements Enrico Bramantyo Hady¹, Dana Sulisty Kusumo²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹clementsceb@students.telkomuniversity.ac.id, ²danakusumo@telkomuniversity.ac.id

Abstrak

Pengembangan software berbasis iteratif sudah sering diterapkan dalam dunia kerja. Proyek software pada era modern menuntut produk disampaikan secara cepat dalam setiap sprint pengembangannya. Maka dari itu, eksekusi dari suatu sprint memerlukan pemantauan dan kemampuan untuk memberikan produk dengan kualitas tinggi seiring dengan tahapan pengembangan antar sprint. Penelitian ini bertujuan untuk memberikan dukungan bagi penanggung jawab atau pemilik proyek dalam memprediksi kemampuan suatu produk dalam suatu sprint. Metode yang diusulkan untuk mencapai tujuan ini adalah membangun sebuah prediction model menggunakan sejumlah fitur berupa karakteristik dari dataset yang memuat sprint-sprint dari software project. Model yang diusulkan dibentuk menggunakan metode utama Random Forest Regressor dengan metode pembanding berupa KNN(K-Nearest Neighbours) dan Decision Tree Regressor. Hasil pengujian dari prediction model yang diusulkan menunjukkan bahwa dibandingkan dengan KNN dan Decision Tree, Random Forest Regressor memberikan performa terbaik melalui hasil prediksinya yang stabil pada setiap tahap kemajuan dari software project yang diuji.

Kata Kunci : *prediction model, software development, sprint*

Abstract

Iterative-based software development has been frequently implemented in working environment. A modern era software project demands that the product is delivered on every iteration development. Hence, the execution of an iteration requires supervision and capabilities to deliver a high quality product alongside the iteration's development. This research's purpose is to give support for a project's supervisor or owner in predicting a product's capability in an iteration. The proposed method for this purpose is to build a prediction model utilizing a number of features in a form of characteristics from a dataset containing software project iterations. The proposed model is built using Random Forest Regressor as a main method with KNN(K-Nearest Neighbours), and Decision Tree Regressor being the comparison methods. Testing results show that compared to KNN and Decision Tree, Random Forest Regressor yields the best performance through its stable results on every progress stages of all tested software projects.

Keywords: *prediction model, software development, sprint*

1. Pendahuluan

Latar Belakang

Pengembangan *software* secara modern seringkali didasarkan pada pendekatan berbasis iteratif di mana *software* dikembangkan dan diuji dalam siklus berulang [1]. pendekatan ini menguntungkan *developer* karena produk akan selalu mendapat masukan dari hasil sebelumnya yang sudah dipelajari. Ini merupakan salah satu penerapan *sprint* [2], *sprint* merupakan hasil dari *Scrum* [3] [4] [5] yang merupakan salah satu metode *software development* terpopuler.

Tetapi, selalu terdapat ketidakpastian [6] dalam *software development*. Dikarenakan sifat pengerjaannya yang dinamis, melalui perubahan dan masukan yang konstan terhadap sprint, perubahan terhadap kebutuhan produk dapat terjadi. Beberapa contoh lain dari ketidakpastian yang dialami adalah kemungkinan *sprint delivery* yang terlambat dan penggunaan biaya berlebih [7] dalam pengembangan *software project*. Untuk menangani ini, perencanaan matang, pemantauan *progress*, dan interaksi konstan mengenai identifikasi resiko proyek [7] [8] dengan tim *Scrum* [9] serta *customer* merupakan hal penting dalam pengembangan *software* berbasis iteratif.

Penelitian ini berfokus kepada kemampuan *prediction model* untuk memprediksi tingkat *sprint delivery* dalam bentuk performa setiap *software project* pada interval tertentu. Pemberian prediksi didasarkan terhadap fitur atau masalah dalam suatu sprint yang mencakup beberapa hal seperti durasi kerja, change log [7], bobot pengerjaan dan status suatu *sprint* [7]. Dengan mempelajari karakteristik fitur yang mampu mempengaruhi performa [3] dari setiap *sprint*, sebuah *prediction model* dibangun untuk memprediksikan apakah pengerjaan suatu *software project* mampu mencapai target pengerjaan yang ditentukan.

Dengan itu, penelitian ini berfokus untuk menguji apakah target yang diberikan dapat dicapai pada akhir setiap tahap kemajuan *software project*. Melalui ini, penanggung jawab atau pemilik proyek dapat mengetahui seberapa besar performa pengembangan proyek pada setiap tahap kemajuannya. Dengan mengetahui tingkat performa pada setiap interval, penanggung jawab atau pemilik dapat memperkirakan kualitas yang dapat dicapai pada hasil *software project* secara keseluruhan.

Topik dan Batasannya

Topik permasalahan yang diangkat dalam penelitian ini adalah menguji performa *prediction model* dalam memprediksi *sprint delivery* dari sprint yang terdapat dalam sebuah *software project*. Serta membandingkan performa *prediction model* yang diusulkan pada beberapa *software project* lainnya. Adapun batasan-batasan dalam penelitian ini di antaranya, dataset yang digunakan adalah dataset *iterations* dan dataset *issues* dari *software projects* yang digunakan dalam penelitian ini. Metode utama yang digunakan untuk membangun *prediction model* adalah *Random Forest Regressor*. Metode pembanding yang digunakan dalam penelitian ini adalah *K-Nearest Neighbours* dan *Decision Tree Regressor*. Metode pembanding dipilih untuk menguji seberapa besar pengaruh dari efektivitas algoritma dan dimensi data terhadap tingkat akurasi model.

Tujuan

Tujuan dari penelitian ini adalah untuk mengimplementasikan sistem prediksi performa *sprint delivery* dalam *iterative-based software development*. Tidak hanya itu, penelitian ini juga akan mengevaluasi performa *prediction model* dalam memprediksi tingkat *sprint delivery* dari beberapa *software project* yang diuji.

Organisasi Tulisan

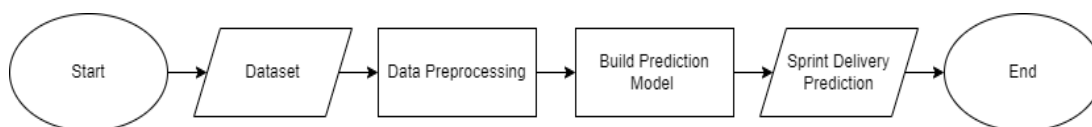
Penelitian ini disusun dengan urutan sebagai berikut: Bab II membahas studi terkait penelitian. Bab III membahas metode penelitian yang digunakan. Bab IV membahas hasil penelitian. Serta Bab V membahas kesimpulan penelitian serta *future works*.

2. Studi Terkait

Studi kami didasarkan pada studi serupa mengenai proyek pengembangan perangkat lunak. Referensi kami adalah studi dalam proyek perangkat lunak mengenai estimasi upaya dan durasi proyek [10], estimasi story points [11], dan prediksi waktu bug fixing [12]. Ada juga prediksi *delivery capabilities* [7]c yang memanfaatkan *directed acyclic graph* yang dibentuk menggunakan *dependencies* dari satu atau lebih *issues* yang menghalangi penyelesaian *issue* lain dalam *sprint*.

Namun, tidak seperti penelitian yang disebutkan sebelumnya mengenai prediksi *delivery capability* [7], penelitian kami tidak memanfaatkan *issue dependencies*. Batasan ini disebabkan karena kendala yang ditemui baik pada waktu penelitian maupun tidak adanya *issue dependencies* pada dataset yang diperoleh [7]. Sebagai gantinya, kami membangun model prediksi yang lebih sederhana menggunakan parameter berupa fitur yang tersedia dengan nilai korelasi tertinggi. Pembangunan model yang lebih sederhana dilakukan dengan pemikiran untuk memberikan alternatif pilihan dengan memberikan metode lain tentang bagaimana memprediksi tingkat performa suatu *software project*.

3. Sistem yang Dibangun



Gambar 1. Alur Kerja Sistem

Alur kerja dari sistem prediksi yang dibangun dimulai dengan melakukan *preprocessing* data dengan cara menghapus kolom yang tidak dibutuhkan. Kemudian dilakukan *encoding* data kategorikal menjadi data numerik. Lalu dilakukan agregasi antara dataset *iteration* dengan dataset *issues*. Data yang telah dibersihkan dan dipersiapkan kemudian akan digunakan untuk membangun *prediction model*.

Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset yang dikembangkan oleh [7], [8], [13], [11] dalam studi dan evaluasi yang bersangkutan. Dalam dataset ini terdapat 3 kelompok data, yaitu dataset mengenai *agile*