

# 1. Pendahuluan

## Latar Belakang

*People Analytics* kini menjadi instrument kunci dalam manajemen kemampuan [15]. Analitik ini berupa gabungan dari penambangan data (*data mining*) dengan *business analytics*, yang kemudian diaplikasikan pada data *human resource* [9]. *Data science* dan analitik data membantu organisasi dengan menyediakan analitik yang deskriptif, prediktif, dan preskriptif [15]. *People Analytics* juga digunakan sebagai solusi baru untuk mengambil keputusan berdasarkan fakta aktual, yang dapat meningkatkan kinerja organisasi [12]. Perusahaan kini dapat mendeteksi dan mengidentifikasi orang seperti apa yang ingin mereka pekerjakan dan apakah calon karyawan mereka dapat meningkatkan kinerja, mencapai tujuan, dan berinovasi. Oleh karena itu, penelitian yang lebih mendalam tentang *people analytics* sangat diperlukan. Hal ini menguntungkan tidak hanya bagi perusahaan yang aktif merekrut talenta baru, tetapi juga bagi para lulusan baru yang akan segera memasuki dunia kerja profesional. Dengan penelitian ini, lulusan baru mengetahui faktor-faktor apa saja yang membantu kesiapan kerja mereka dan strategi untuk memenuhi kriteria perusahaan yang mereka impikan. Semua orang berlomba-lomba untuk mendapatkan pekerjaan, dan itu tidak mudah, apalagi dengan faktor eksternal yang tidak bisa dihindari seperti pandemi. Kesiapan mental, *soft*, dan *hard skills* sangat dibutuhkan bagi lulusan baru untuk beradaptasi dengan lingkungan kerjanya. Kesiapan kerja merupakan suatu keharusan bagi mahasiswa yang saat ini sedang mempersiapkan diri untuk memasuki dunia kerja profesional.

Dalam beberapa studi sebelumnya yang terkait dengan pendekatan *people analytics*, perlu dibangun model klasifikasi yang dapat memprediksi faktor-faktor yang mungkin dapat membantu manajemen bakat dan sumber daya manusia menggunakan pembelajaran mesin. Penelitian ini membangun model untuk memprediksi kesiapan kerja mahasiswa Telkom University dengan menganalisis kinerja mahasiswa selama masa studi aktif menggunakan Multinomial Logistic Regression. Penelitian ini juga menggunakan Random Forest sebagai pembanding untuk mengetahui model mana yang memperoleh nilai akurasi yang lebih baik. Multinomial Logistic Regression telah digunakan dalam penelitian dengan data kategoris namun menghasilkan skor akurasi yang relatif kurang baik untuk dataset kecil, seperti yang ditunjukkan dalam penelitian [14]. Kemudian dibuktikan bahwa Random Forest dapat menghasilkan performa model yang cukup baik dengan tingkat kesalahan rata-rata 0%. Tetapi seleksi fitur yang digunakan pada penelitian tersebut adalah ANOVA, yang digunakan dalam penelitian dengan variabel bebas bersifat kategoris dan variabel terikat bersifat numerik. Penelitian ini menggunakan *Chi-square Test* sebagai seleksi fitur karena variabel bebas maupun variabel terikat dalam penelitian ini bertipe kategoris. Penelitian ini kemudian melakukan eksperimen dengan SMOTE-ENN untuk menyeimbangkan data untuk membuktikan apakah data yang tidak seimbang berpengaruh signifikan terhadap skor akurasi. Pada penelitian [13], SMOTE-ENN dapat meningkatkan skor akurasi sekitar 3%. Selain itu, penelitian ini juga bertujuan untuk mengetahui faktor dan kompetensi yang mendukung kesiapan kerja mahasiswa dan mendapatkan pekerjaan tetap setelah lulus.

## Topik dan Batasannya

Pada penelitian ini, Penulis akan membangun model klasifikasi untuk memprediksi kesiapan kerja Mahasiswa Telkom University. Model sistem akan terdiri dari *preprocessing*, seleksi fitur dengan *Chi-square Test*, klasifikasi dengan Multinomial Logistic Regression dan Random Forest, serta validasi dengan *confusion matrix*. Kemudian batasan masalah untuk penelitian ini adalah terdapat 3 kelas kategori untuk mengklasifikasikan hasil prediksi, yaitu Sebelum Lulus, Dibawah Rerata, dan Diatas Rerata. Kemudian data yang digunakan adalah Data Tracer Study Telkom University dari tahun 2015 hingga 2020 yang berjumlah lebih dari 9000 baris dan data IPK Mahasiswa yang berjumlah lebih dari 28000 baris.

## Tujuan

Tujuan dari penelitian ini adalah untuk memprediksi kesiapan kerja mahasiswa Telkom University dengan menganalisis kinerja mahasiswa selama masa studi aktif menggunakan Multinomial Logistic Regression dan Random Forest. Penelitian ini juga bertujuan untuk mengetahui apakah penanganan data yang tidak seimbang dengan SMOTE-ENN dapat meningkatkan nilai akurasi model. Selain itu, penelitian ini juga bertujuan untuk mengetahui faktor dan kompetensi yang mendukung kesiapan kerja mahasiswa.

## Organisasi Tulisan

Pada Bab 2, akan dijelaskan mengenai studi terkait yang mendukung dengan penelitian ini. Kemudian pada Bab 3 akan dijelaskan rancangan dan model yang digunakan pada penelitian ini. Bab 4 akan menjelaskan hasil dari sistem yang dibangun serta menganalisis hasil pengujian terhadap sistem. Terakhir, pada Bab 5 akan menjelaskan kesimpulan pada penelitian ini serta saran yang dapat membantu penelitian selanjutnya mengenai topik yang sama.