**BAB 3**

**Metodologi**

* 1. **Analisis Masalah**

Setiap manusia memiliki kepribadiannya masing-masing. Kepribadian ini menjadi ciri unik yang menjadi gambaran seseorang tersebut di dalam kehidupannya. Kepribadian dapat membedakan orang tersebut dalam berperilaku atau memecahkan suatu permasalahan. Saat ini, kepribadian juga menjadi sesuatu yang sangat penting untuk dijadikan pertimbangan dalam perekrutan tenaga kerja. Di sisi lain, kepribadian juga dapat menjadi faktor suatu hubungan dan relasi terhadap orang lain yang cocok kepribadiannya satu sama lain dan masih banyak lagi pengaruh kepribadian terhadap kehidupan seseorang.

Untuk itu, banyak sekali cara yang dapat dilakukan oleh seseorang untuk mengetahui kepribadiannya sendiri. Mulai dari mengisi kuesioner kepribadian atau aplikasi yang beredar di sosial media dan internet. Salah satu model kepribadian yang paling sering digunakan adalah *The Big Five Model Personality*. Model kepribadian ini terbagi menjadi 5 bagian utama yaitu, *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness* dan *Neuroticism*. Cara mendapatkan hasil kepribadian dari model ini biasa dengan mengisi beberapa pertanyaan mengenai diri kita sendiri. Namun pada kenyataannya, hasil dari kepribadian ini biasa tidak tepat dan dapat dimanipulasi oleh orang yang mengisinya. Faktor-faktor tersebut biasa berupa untuk menutupi kepribadian dirinya yang ia anggap buruk atau pertanyaan yang terlalu banyak sehingga menyebabkan responden merasa bosan dan kesulitan untuk menjawab semuanya.

Menurut Pennebaker, Mehl, & Gosling (2006), kepribadian seseorang akan lebih akurat bila dinilai berdasarkan perilakunya sehari-hari. Perilaku seseorang ini biasa hanya dapat dilihat dan dirasakan oleh orang-orang yang berada di sekitarnya. Tetapi, akibat perkembangan teknologi zaman saat ini, dimana orang lebih banyak menghabiskan waktunya di dunia maya atau sosial media, kepribadian seseorang secara tidak langsung dapat dilihat oleh siapa saja.

Walaupun penelitian mengenai sistem prediksi kepribadian dari sosial media telah banyak dilakukan. Namun, masih sebagian kecil yang menggunakan sosial media Facebook karena tidak tersedianya *Public API* oleh Facebook untuk mendapatkan dataset dari *user*. Beberapa penelitian yang menggunakan sosial media Facebook juga masih memiliki akurasi yang tergolong rendah.

* 1. **Usulan Pemecahan Masalah**

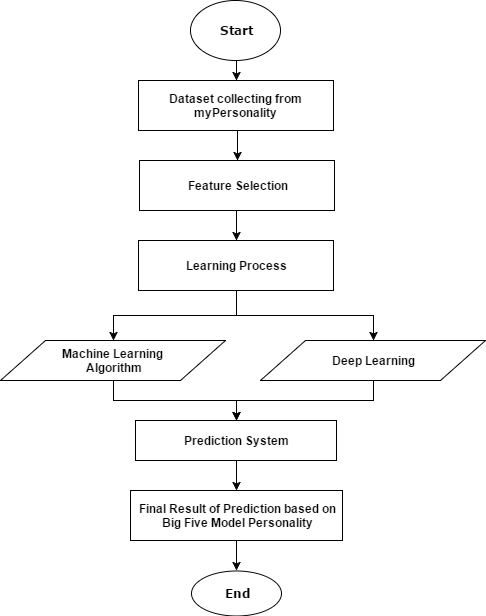
Dari analisis permasalahan tentang penelitian sistem prediksi kepribadian, peneliti berencana mengembangkan sistem prediksi kepribadian dengan akurasi yang lebih tinggi dari penelitian sebelumnya yang menggunakan sosial media Facebook. Masalah dalam sistem prediksi ini yang paling penting adalah tingkat akurasi. Untuk itu peneliti berencana mencoba berbagai metode sebagai berikut:

1. Mencoba berbagai algoritma *Machine Learning* untuk *training* dan membandingkan akurasi yang paling tinggi untuk setiap *traits* dari *Big Five*.

2. Melakukan perbandingan berbagai *feature pre-defined* (*closed vocabulary*) yang paling akurat untuk sistem prediksi.

3. Mencoba menerapkan *Deep Learning* untuk training sistem dengan metode *open* *vocabulary*.

* 1. **Kerangka Berpikir**

****

**Gambar 3.1 Kerangka Berpikir**

**3.3.1 *Dataset Collecting from myPersonality***

Sosial media Facebook tidak memberikan *Public API* untuk mendapatkan data dari *user*. Facebook hanya menyediakan *API* untuk mendapatkan data dari account kita sendiri atau account lain selama kita memiliki authentication code dari account tersebut. Maka untuk mendapatkan dataset yang lebih besar, kami menggunakan data user Facebook dari myPersonality (Kosinki, 2015). myPersonality adalah aplikasi Facebook yang dikembangkan oleh Michal Kosinki dimana user Facebook dapat menggunakan aplikasi itu untuk mendapatkan hasil kepribadian mereka berdasarkan model kepribadian Big Five. Data yang disediakan secara terbuka hanya berupa data 250 user dengan kurang lebih 10.000 status. Data-data ini telah ditambahkan dengan beberapa detail seperti waktu, social network dari user dan jenis kepribadiannya.

**Tabel 3.1 Distribusi jenis kepribadian berdasarkan**

**250 user dalam dataset (Kosinki, 2015)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Value** | **cOPN** | **cCON** | **cEXT** | **cAGR** | **cNEU** |
| Yes | 176 | 130 | 96 | 134 | 99 |
| No | 74 | 120 | 154 | 116 | 151 |

**3.3.2 *Feature Selection***

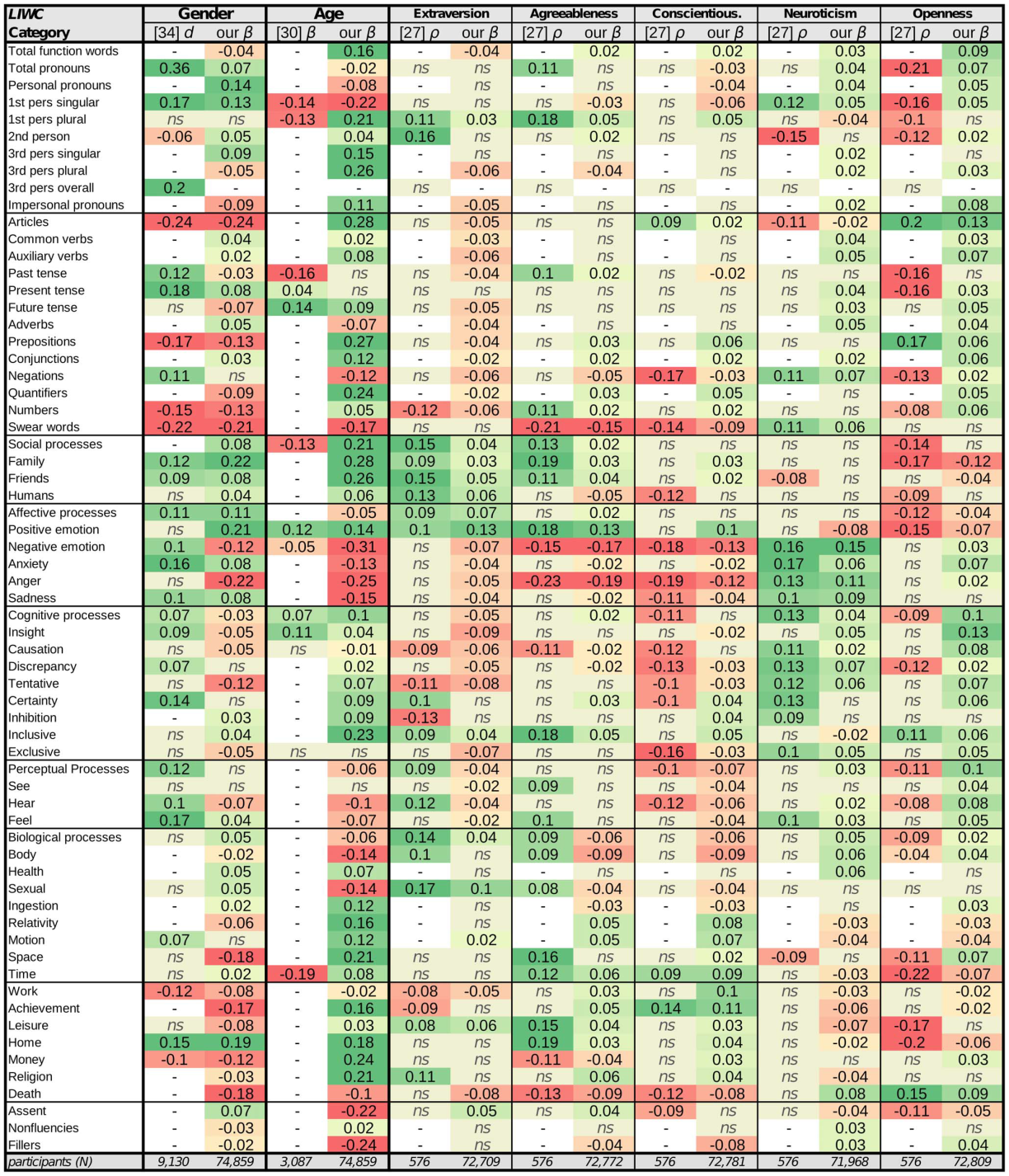
Dalam pengembangan sistem dan setelah melalui beberapa pencarian mengenai fitur yang dapat digunakan dalam melakukan proses olah bahasa. Beberapa fitur di bawah ini merupakan fitur yang kami coba dan kemudian kami bandingkan satu sama lain untuk dilihat keakuratan dan fungsionalitasnya. Berikut beberapa fitur yang ada dalam penelitian ini:

1. **LIWC**

Fitur linguistik pertama yang kita terapkan dalam penelitian adalah fitur LIWC (*Linguistic Inquiry and Word Count*). Fitur LIWC ini dapat dikatakan sebagai fitur yang paling umum dan paling sering digunakan dalam penelitian NLP karena telah lama dikembangkan dan diperbaharui hingga saat ini. Sebagai awal mula pembangunan sistem penelitian kami mencoba untuk mengumpulkan seluruh status dari user untuk kemudian diibagi per masing-masing user.

Setelah didapatkan susunan status per user. Kemudian kami menggunakan API yang disediakan LIWC melalui Receptiviti untuk mendapatkan hasil dari nilai fitur dengan status dari user.

Tabel 3.2 Korelasi Kategori LIWC dengan jenis kelamin, umur dan *Big Five model personality* (Schwartz, et al., 2013)



1. **Splice**

Fitur linguistik kedua yang kami gunakan adalah SPLICE (*Structured Programming for Linguistic Cue Extraction).* Fitur ini masih cukup baru dan masih jarang digunakan, tetapi setelah melalui proses research paper dan literature. Peneliti merasa fitur ini cukup akurat dan lengkap. Proses yang dilakukan mirip dengan ketika peneliti melakukan korelasi linguistik dengan fitur LIWC. Setelah status dipilah berdasarkan masing-masing user. Kemudian API dari SPLICE dipanggil dan digunakan untuk menghasilkan nilai dari fitur-fitur yang ada di SPLICE.

1. **Social Network Analysis (SNA) Features**

Fitur selanjutnya merupakan fitur *Social Networks Analysis* dari user. *Social Networks Analysis* adalah suatu alat atau studi yang memetakan hubungan pengetahuan yang penting antar individu (Pryke, 2004). SNA dikembangkan untuk memahami hubungan –hubungan (*ties/edge*) dari aktor-aktor (*nodes/points*) yang ada dalam sebuah sistem dengan 2 fokus, yaitu aktor-aktor dan hubungan antar actor dalam konteks sosial tertentu. SNA sering diimplementasikan untuk mengidentifikasi arus informasi. Secara teori dengan mengidentifikasi arus informasi bisa membantu meningkatkan strategi yang bisa memacu para aktor untuk berbagi informasi daripada harus menciptakan strategi yang baru (Serrat, 2009).

Fitur dari SNA yang digunakan dalam penelitian ini terbagi atas 7 fitur yaitu:

* *Network size*
* *Betweeenness*
* *nBetweenness*
* *Density*
* *Brokerage*
* *nBrokerage*
* *Transitivity*

1. **Open Vocabulary**

Metode Open Vocabulary ini berbeda dengan fitur-fitur lainnya dimana semua nilai dari fitur di define di awal sebelum proses training dilakukan. Metode Open Vocabulary tidak memiliki jumlah fitur pasti, tetapi fitur didapatkan dengan menelusuri dataset yang kita gunakan. Kemudian, kata-kata dari dataset yang kita kumpulkan dibagi dan dihitung jumlahnya (word count). Setiap kata itulah yang dikembangkan menjadi fitur dan dibandingkan dengan Big Five Model Personality.

Berdasarkan penelitian sebelumnya, masih banyak pertimbangan dalam NLP untuk menentukan metode manakah yang lebih baik. Apakah metode Closed Vocabulary (LIWC, SPLICE, MRC, dan lainnya) atau dengan metode Open Vocabulary. Untuk itu dalam penelitian ini kami juga menerapkan metode Open Vocabulary untuk mendapatkan hasil akurasi prediksi dan dibandingkan dengan metode lainnya.

**3.3.3 *Learning Process***

Proses learning dibagi menjadi dua tahapan, yaitu dengan menggunakan machine learning algorithm dan percobaan penggunaan deep learning.

**1. Machine learning**

Setelah melalui tahap seleksi fitur, hasil-hasil tersebut kemudian dilanjutkan ke dalam proses learning dengan menerapkan beberapa algoritma classifier yaitu SVM, gradient boosting, Linear Regression dan lainnya.

**2. Deep learning**

Setelah mendapatkan hasil dari proses learning tersebut. Peneliti kemudian mencoba kembali menggunakan proses Deep Learning dengan beberapa arsitektur seperti MLP, CNN, dan LSTM. Hasil dari penggunaan fitur yang berbeda dan proses learning yang berbeda akan terus dibandingkan akurasinya untuk mendapatkan sistem prediksi yang paling akurat.

**3.3.4 *Prediction System***

Pada tahap ini data training dari sistem yang telah melewati proses learning akan dibandingkan dengan data testing. Melalui tahap ini, akan didapatkan prediksi terhadap data testing tersebut. Peneliti akan melihat tingkat akurasi dari prediksi tersebut untuk kemudian dikembangkan secara maksimal untuk mencapai tingkat akurasi yang paling maksimal berdasarkan 3 faktor paling menentukan dalam hasil sistem prediksi ini yang disebut oleh peneliti dengan FAL (*Feature extraction*, *Algorithm* *classifier*, and *Learning* *process*).

Hasil klasifikasi dari *training* ini akan menampilkan boolean value untuk setiap *trait* kepribadian dari *Big Five Model Personality*. Nilai 1 untuk *user* yang dianggap memiliki persentase tinggi di *trait* tersebut dan nilai 0 untuk *user* yang dianggap memiliki persentase rendah di *trait* tersebut.

**3.3.5 *Final result of Prediction based on Big Five model***

Tahap akhir dari penelitian akan berwujud sebuah user interface untuk menampilkan secara langsung hasil prediksi dari seorang user yang telah melakukan log in ke dalam Facebook sehingga terhubung ke API Facebook untuk mendapatkan data-data yang dibutuhkan dari user tersebut khususnya status. Hasil prediksi dalam interface akan menunjukkan kepribadian Big Five Model Personality dari user tersebut disertai dengan fitur-fitur dan informasi lainnya yang dikembangkan lebih lanjut.

* 1. **Rancangan Layar**

Hasil akhir dari aplikasi akan berupa sebuah aplikasi berbasis *web* yang memungkinkan pengguna untuk mencoba langsung sistem prediksi yang telah kami kembangkan dan mendapatkan hasil dari kepribadian mereka dengan terhubung ke akun Facebook mereka. Berikut adalah perancangan layar aplikasi yang akan dibuat:

**3.4.1 Rancangan *Homepage* aplikasi**