

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Menurut WHO, peran penting kesadaran akan kesehatan mental mendukung tujuan perkembangan secara global. Depresi adalah salah satu faktor penting dalam permasalahan kesehatan mental. Bunuh diri adalah resiko kematian kedua yang terjadi kepada sebagian masyarakat dari umur 15 - 29 tahun. Pasien dengan kesehatan mental parah dapat mengalami kematian prematur dikarenakan kondisi fisik yang sebenarnya dapat dicegah. Penelitian akan kesehatan mental belakangan ini juga meningkat. Hal ini dipicu oleh stigma negatif masyarakat terhadap gangguan kesehatan mental pada umumnya. Penderita gangguan kesehatan mental dapat merasa malu dan akhirnya tidak mencari bantuan atau perawatan medis yang layak. Untuk itulah dibuat penelitian demi penelitian akan kesehatan mental untuk meningkatkan kewaspadaan masyarakat akan kesehatan mental terutama pada bidang teknologi dan IT sebagai bidang pekerjaan yang saat ini banyak diburu oleh perusahaan.

Pada penelitian ini, tim peneliti memutuskan untuk membangun model untuk memprediksi positif atau tidaknya pekerja IT memiliki gangguan kesehatan mental. Data yang dipakai untuk melatih model diambil dari hasil survey OSMI (Open Sourcing Mental Illness, LTD) yang dimulai pada tahun 2016 dan penambahan data terakhir dilakukan pada dua tahun yang lalu. Survey tersebut dilakukan untuk mengetahui bagaimana pandangan masyarakat terhadap kesehatan mental khususnya pekerja yang bekerja di bidang teknologi atau IT. Hasil survey tersebut ditargetkan untuk meningkatkan kesadaran akan pentingnya kesehatan mental dan menciptakan suasana tempat kerja yang nyaman bagi mereka yang memiliki gangguan kesehatan mental. Dataset yang diperoleh akan dikembangkan dan dilakukan normalisasi terhadap data yang kemudian akan dilakukan seleksi fitur secara unik dan visualisasi data fitur yang terseleksi.

Analisa data sekarang telah banyak memanfaatkan konsep model *Machine Learning* dan *Deep Learning*. Model yang dibangun pada penelitian ini akan menggunakan algoritma SVM (Support Vector Machine) dan FCNN (Fully

Connected Neural Network). Metode algoritma yang digunakan ada dua sebagai perbandingan dikarenakan dalam *data science* belum tentu model dengan metode neural network (*Deep Learning*) dapat mengungguli performa SVM yang merupakan salah satu algoritma *Machine Learning* dasar. Hasil akurasi pada akhirnya hanya dapat ditentukan dari bagaimana model yang dibangun dengan metode yang digunakan dapat menghasilkan grafik prediksi yang paling menyerupai grafik data sebenarnya.

SVM dalam hal ini digunakan sebagai algoritma *Machine Learning* yang sering digunakan sebagai acuan valid untuk perbandingan performa akan FCNN yang dibangun. Algoritma FCNN yang dibangun sendiri merupakan bagian dari *Artificial Neural Network* (ANN) dengan perkembangan bahwa lapisannya saling terhubung atau yang lebih dikenal sebagai Dense Neural Network (DNN). Kedua desain algoritma tersebut kemudian akan dilakukan penyesuaian dan performa akhirnya akan dibandingkan untuk menentukan metode yang paling tepat untuk diterapkan pada dataset yang digunakan. Hasil penelitian ini diharapkan dapat menghasilkan metode analisa yang terbaik terhadap prediksi gangguan kesehatan mental.

1.2. Rumusan Masalah

Adapun rumusan masalah yaitu: Bagaimana penerapan algoritma Fully Connected Neural Network dan Support Vector Machine dalam memprediksi gangguan kesehatan mental dengan menggunakan hasil analisa data terhadap survey kesehatan mental pada para pekerja di bidang teknologi?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Adapun tujuan dari penelitian ini adalah untuk meningkatkan kewaspadaan masyarakat (*public awareness*) terhadap kesehatan mental dan menunjukkan bahwa model yang dibangun mampu memprediksi gangguan kesehatan mental dengan menggunakan hasil analisa data dari tim peneliti terhadap hasil survey kesehatan mental pada para pekerja di bagian teknologi, serta menentukan apakah teknik *Deep Learning* (FCNN) dapat mengungguli teknik *Machine Learning* (SVM).

1.3.2. Manfaat

Manfaat dari penelitian ini adalah memberikan analisa data terhadap hasil survey dan membangun model yang dapat digunakan untuk memprediksi gangguan kesehatan mental pekerja teknologi. Model yang dibangun diharapkan dapat membantu dalam menciptakan kondisi masyarakat yang sadar akan pentingnya kesehatan mental dan sebagai bahan acuan pembelajaran untuk teknik prediksi dan teknik analisa data terhadap kesehatan mental pada masa yang akan datang. Analisa data yang didapatkan diharapkan mampu menyajikan hasil informasi dan isi penting dari hasil survey yang didapatkan sehingga informasi lebih mudah tersampaikan dan data yang siap untuk diimplementasikan.

1.4. Batasan Masalah

Dalam penelitian ini, terdapat batasan masalah dari permasalahan yang ada, yaitu :

1. Dataset terdiri dari 1467 entri data yang merupakan representatif data hasil survey yang dilakukan terhadap pekerja teknologi. Dataset dilaksanakan oleh OSMI (Open Sourcing Mental Illness, LTD) yang dapat diakses secara online melalui <https://www.kaggle.com/osmi/mental-health-in-tech-2016>.
2. Model yang dibangun menggunakan algoritma Fully Connected Neural Network dan Support Vector Machine.
3. Bahasa pemrograman yang digunakan adalah bahasa Python dengan memanfaatkan *library* yang telah ada. SVM akan dibantu dengan menggunakan bantuan *library sklearn* sedangkan FCNN akan dibangun dengan memanfaatkan *library keras* dan *tensorflow*.
4. Hasil penelitian difokuskan dengan ditentukan ada atau tidaknya para pekerja di bagian teknologi pada saat ini memiliki gangguan kesehatan sebagaimana data dependen (target) klasifikasi yang dimana jawabannya terbagi menjadi tiga yaitu: “Yes”, “No”, dan “Maybe”. Validasi terhadap jawaban tersebut didukung oleh pertanyaan survey yang menanyakan “*If yes, what condition(s) have you been diagnosed with?*” dan “*If maybe, what condition(s) do you believe you have?*” yang akan direspon dengan baik oleh peserta survey. Data

independen (*feature*) terdiri dari 42 kolom hasil seleksi fitur dengan mengeliminasi kolom yang mempunyai banyak *missing values*. Seluruh kolom akan dilakukan tahap *preprocessing* terlebih dahulu dari penamaan ulang hingga normalisasi data.

1.5. Keterbaruan

Penelitian serupa sebelumnya telah dilakukan dengan judul “*Predicting mental health problems in adolescence using machine learning techniques*”. Tujuan penelitian tersebut adalah membangun model yang dapat memprediksi permasalahan kesehatan mental pada remaja menengah keatas dan menentukan apakah teknik *machine learning* (*random forest*, *svm*, *neural network*, dan *XGBoost*) dapat mengungguli *logistic regression*. Hasil penelitian tersebut yang peforma di ukur dengan teknik AUC (*Area rea under the ROC Curve*) walaupun menghasilkan hasil yang bervariasi menunjukkan sedikit keunggulan pada model *random forest* dengan AUC = 0.739,95% CI 0.708-0.769) diikuti *svm* dengan AUC = 0.735,95% CI 0.707-0.764). Kesimpulan dari penelitian tersebut menunjukkan model tersebut tidak layak digunakan untuk keperluan medis, tetapi penelitian tersebut menghasilkan dasar penelitian yang baik bagi penelitian yang akan datang (yang oleh tim peneliti tersebut disarankan untuk tidak memakai *logistic regression* kembali sebagai dasar acuan)[1].

Penelitian selanjutnya yang patut untuk disinggung adalah penelitian dengan judul “*Predicting Mental Conditions Based on “History of Present Illness” in Psychiatric Notes with Deep Neural Networks*”. Penelitian tersebut merupakan tahap ketiga dari *roadmap* penelitian yang dilakukan tim tersebut yang tugasnya terbagi dan menawarkan kasus penggunaan data yang baru dari dataset yang dipakai. Dataset yang dipakai terdiri dari 1000 entri data yang selanjutnya dilakukan *preprocessing* dan normalisasi data hingga tersisa 986 data siap pakai. Data dependen yang digunakan memiliki dua respon, yakni: “*Yes*” dan “*No*”. Model dibuat dengan algoritma CNN (*Convolutional Neural Networks*) dan ReHAN (*Recurrent Neural Networks with Hierarchical Attention*) yang dibangun terpisah. Eksperimen dilakukan dengan membandingkan kedua peforma dari algoritma tersebut

menggunakan metode *mean micro-F1 score* yang menunjukkan data dependen yang dipakai merupakan *multi-label binary* (lebih dari satu kolom data dependen dengan keseluruhan kolom bertipe data Boolean). Model dengan algoritma CNN memberikan hasil peforma terbaik dengan 63.144% *mean micro-F score* dan ReHAN 61.904% *mean micro-F1 score*[2].

Pada penelitian yang ditulis peneliti, berbeda dengan penelitian yang sebelumnya, tim peneliti menggunakan tingkat akurasi sebagai ukuran dan mencantumkan dimensi matrix dari model algoritma sebagai validasi untuk memastikan akurasi yang didapatkan merupakan hasil prediksi yang seimbang dan valid. Model yang dipakai dalam penelitian peneliti memiliki data dependen dengan tiga respon, yakni: “Yes”, “Maybe”, dan “No” sehingga tergolong kedalam *multi-class classification*. Peforma yang diraih oleh model juga pastinya diekspektasikan dapat meraih angka diatas 60%.

Penelitian lainnya yang dijadikan bahan referensi diantaranya: Penelitian yang menggunakan FCNN namun dengan penambahan metode *Long short-term memory* (LSTM-FC) untuk memprediksi konsentrasi PM2.5[3]; Penelitian tentang prediksi kesehatan mental mahasiswa dengan metode modal psikologis (pendekatan secara psikologi)[4]. Mengacu pada penelitian yang dijabarkan diatas tim peneliti berharap dapat melakukan analisa data dengan baik dan menghasilkan peforma model yang baik dan menentukan algoritma terbaik yang tepat untuk diimplementasikan pada hasil analisa dataset sehingga penelitian ini dapat bermanfaat untuk keperluan penelitian berikutnya maupun dalam hal memberi gambaran baru terhadap kesehatan mental sehingga berkurangnya stigma negatif dalam masyarakat mengenai kesehatan mental khususnya bagi pekerja teknologi.