

Purwarupa Sistem Klasifikasi Legalitas Investasi Berbasis Algoritma *Bidirectional Long Short Term Memory*

Muhammad Navi Nugraha¹, Muhammad Arrofiq^{1*}

¹Departemen Teknik Elektro dan Informatika, Sekolah Vokasi, Universitas Gadjah Mada;
navinugraha15@gmail.com

*Korespondensi: rofiq@ugm.ac.id;

Abstract – *Illegal investment has become a serious problem in Indonesia, causing significant financial losses for investors and the economy. Manual verification of investment company legality has proven ineffective and time-consuming. Therefore, this research aims to enhance the effectiveness of investment legality verification by utilizing machine learning technology for automated prediction of investment messages. The study employs text classification methodology to categorize Indonesian investment messages from social media platforms, particularly Telegram, into legal or illegal categories. The research objective is to design a prototype investment legality detection system. The dataset used in this study is sourced from Telegram, comprising two variants with 2996 and 210 data points, respectively, covering legal and illegal investment messages. The research process involves data collection, preprocessing, classification model training, evaluation, and the development of a web application for investment message detection. Two algorithms, BiLSTM and LSTM, are implemented and compared in this research. The evaluation considers dataset size, dataset split ratio, utilization of BiLSTM and LSTM algorithms, confusion matrix analysis, and the web application. The findings reveal that the text classification model with BiLSTM algorithm and a dataset size of 2996 and 70:30 split ratio achieves the highest performance compared to other variants. The BiLSTM model achieves an accuracy of 96%, precision of 98%, and recall of 93%. The web application successfully detects investment messages using the top-performing model and operates according to the designated design.*

Keywords – *Illegal Investments, Text classification, LSTM, Telegram, Natural Language Processing, Ruby on Rails.*

Intisari – Investasi ilegal telah menjadi permasalahan serius di Indonesia, yang mengakibatkan kerugian finansial yang signifikan bagi investor dan perekonomian. Pengecekan legalitas perusahaan investasi secara manual belum efektif dan memakan waktu. Oleh karena itu, penelitian ini memanfaatkan teknologi *machine learning* dengan metode *text classification* untuk melakukan prediksi otomatis terhadap pesan investasi. Metode *text classification* digunakan dalam penelitian ini untuk mengategorikan pesan investasi berbahasa Indonesia dari media sosial, terutama Telegram, menjadi legal atau ilegal. Tujuan penelitian ini adalah merancang purwarupa sistem deteksi legalitas investasi. *Dataset* yang digunakan diperoleh dari media sosial Telegram, terdiri dari dua varian dengan jumlah 2996 data dan 210 data yang mencakup pesan investasi legal dan ilegal. Proses penelitian meliputi pengumpulan data, *preprocessing*, pelatihan model klasifikasi, evaluasi, dan pembuatan aplikasi web untuk deteksi pesan investasi. Dalam penelitian ini, dua algoritma, yaitu BiLSTM dan LSTM, diimplementasikan dan dibandingkan. Pengujian dilakukan dengan mempertimbangkan jumlah data dalam *dataset*, rasio pembagian *dataset*, penggunaan algoritma BiLSTM dan LSTM, *confusion matrix*, serta aplikasi web. Hasil pengujian menunjukkan bahwa model klasifikasi teks dengan algoritma BiLSTM dan jumlah data 2996 dengan rasio pembagian 70:30 memberikan performa tertinggi dibandingkan varian lain. Model dengan algoritma BiLSTM tersebut mampu mencapai akurasi sebesar 96%, presisi 98%, dan *recall* 93%. Aplikasi web berhasil mendeteksi pesan investasi menggunakan model dengan performa tertinggi dan berjalan sesuai dengan rancangan yang telah ditetapkan.

Kata kunci – *Investasi Ilegal, Text classification, LSTM, Telegram, Natural Language Processing, Ruby on Rails.*

I. PENDAHULUAN

Peningkatan jumlah investor di pasar modal Indonesia pada Desember tahun 2022 telah meningkat sebanyak 37,5 persen menjadi 10,3 juta investor dari sebelumnya hanya 7,48 juta investor per Desember 2021 [1]. Hal tersebut menandakan bahwa minat masyarakat Indonesia untuk berinvestasi meningkat dengan didukungnya simplikasi teknologi dalam pembukaan rekening investasi. Peningkatan jumlah investor tersebut menjadi pasar yang menarik untuk para pengusaha mendapatkan profit dengan cara menawarkan berbagai keuntungan investasi di platform masing-masing. Akan tetapi, ada juga peningkatan yang memprihatinkan dari penawaran investasi di Indonesia, yaitu kenaikan skema investasi ilegal di Indonesia yang menjadi isu kritical baik bagi investor maupun pihak berwenang.

Akhir-akhir ini kasus investasi ilegal marak terjadi di Indonesia sehingga menimbulkan kerugian bagi korban.

Investasi disebut ilegal apabila melakukan kegiatan dengan tidak memiliki izin dari regulator jasa keuangan karena tidak memenuhi syarat untuk melakukan kegiatan investasi. Begitu juga melakukan penipuan investasi pada instrumen saham, obligasi, komoditas, kemitraan atau jenis investasi lainnya yang dilakukan dengan janji dan kesepakatan yang menipu [2]. Kasus ini menyebabkan kerugian besar secara finansial bagi para investor maupun ekonomi. Sebagaimana pihak Satgas Waspada Investasi menyebutkan bahwa kerugian keseluruhan masyarakat Indonesia akibat penipuan investasi ilegal selama 5 tahun terakhir telah mencapai 123,5 triliun rupiah [3]. Pendekatan yang digunakan oleh pelaku investasi ilegal untuk menarik calon korban adalah melalui informasi iklan berbasis teks di media sosial dan platform *online*. Menurut survei Populix, dari seribu orang responden, mayoritas sebanyak 68 persen responden tertarik dengan investasi karena memperoleh informasi produk investasi melalui media sosial [4]. Dari kondisi tersebut menunjukkan

bahwa media sosial memiliki pengaruh yang signifikan terhadap minat calon investor untuk menaruh dana mereka dalam investasi ilegal.

Direktorat Jenderal Kekayaan Negara pada tahun 2022 menyarankan upaya pencegahan penipuan investasi ilegal di platform *online* dapat dilakukan dengan pengecekan legalitas perusahaan yang menawarkan investasi. Hendaknya perusahaan investasi sudah mendapatkan izin dan diawasi oleh regulator di sektor jasa keuangan. Salah satu cara untuk melakukan pengecekan legalitas perusahaan investasi secara manual adalah dengan mengakses situs resmi Otoritas Jasa Keuangan (OJK) dan otoritas keuangan lainnya. Namun, cara pengecekan secara manual dinilai kurang efektif karena setiap individu perlu melakukan pengecekan satu per satu legalitas di situs OJK sehingga memerlukan waktu lama bagi calon investor. Salah satu solusi yang lebih efektif adalah dengan menggunakan teknologi *machine learning* yang mampu melakukan prediksi secara otomatis. Metode *machine learning* yang dapat digunakan untuk melakukan klasifikasi berdasarkan legalitas investasi adalah *text classification*. Suatu teknik dari *Natural Language Processing* (NLP) yang dapat menetapkan kategori pada *unstructured text* dari data berformat teks [5]. Penelitian terdahulu tentang *text classification* menunjukkan bahwa dengan menggunakan algoritma Naive Bayes pada *text classification* memiliki hasil akurasi sebesar 82,6% untuk mendeteksi berita *hoax* berbahasa Indonesia secara daring [6]. Penelitian lain oleh Norman dkk. (2018) menambahkan bahwa penggunaan algoritma *Support Vector Machine* (SVM) dan *Natural Language Processing* (NLP) lebih efektif dan akurat dalam memprediksi kategori dari ulasan suatu produk pada *e-commerce*. Pratama [7] melakukan penelitian *text classification* data SMS berbahasa Indonesia untuk deteksi spam dengan mengimplementasikan algoritma *Long-Short Term Memory* (LSTM). Hasilnya menunjukkan bahwa LSTM (akurasi 94%) memiliki akurasi 27% lebih tinggi dibandingkan dengan metode Naive Bayes (akurasi 67%) dan 24% lebih tinggi dari metode K-Nearest Neighbor (KNN) (akurasi 70%). Penelitian serupa pernah dilakukan oleh Nugroho dkk. [8] dimana *text classification* untuk deteksi depresi dan kecemasan pengguna Twitter menggunakan pengembangan dari LSTM, yaitu *Bidirectional LSTM* memiliki tingkat akurasi yang lebih tinggi dibandingkan LSTM. Berdasarkan uraian di atas, maka akan dilakukan perbandingan performa BiLSTM dan LSTM dalam *text classification* pesan investasi. Algoritma yang menunjukkan performa terbaik diharapkan dapat diaplikasikan untuk mendeteksi legalitas investasi dari teks pesan investasi berbahasa Indonesia dan mengategorikannya menjadi ilegal atau legal. Kategorisasi ini dapat digunakan untuk mengidentifikasi investasi ilegal yang berisi informasi palsu atau penawaran keuntungan yang tidak realistis dari media sosial. Hal tersebut seringkali diindikasikan sebagai skema investasi ilegal.

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan sebuah sistem purwarupa yang dapat mengklasifikasikan legalitas investasi dengan

lebih akurat dan mampu mengekstrak informasi lebih lengkap. Sistem purwarupa ini akan mengimplementasikan *text classification* menggunakan algoritma *Bidirectional Long Short-Term Memory* (BiLSTM) dengan menerapkan pustaka algoritma BiLSTM dari Keras. Sistem akan mengambil teks berbahasa Indonesia yang telah diekstraksi oleh model yang dilatih di TensorFlow menggunakan *dataset* yang dikumpulkan dari grup investasi Telegram. Hasil deteksi akan ditampilkan aplikasi web interaktif yang dikembangkan menggunakan Ruby on Rails dan di-*deploy* di fly.io. Penelitian ini juga akan mencakup pengujian performa berupa akurasi, presisi, dan *recall* dari *text classification* menggunakan metode pengujian *confusion matrix*. Dengan memanfaatkan teknologi *machine learning* dengan metode *text classification* dan algoritma BiLSTM, diharapkan dapat membantu mengurangi risiko calon investor terkena investasi ilegal dengan mendeteksi legalitas investasi dan potensi penipuan investasi. Tidak hanya itu, sistem ini diharapkan dapat diimplementasikan dalam bidang pengawasan investasi untuk membantu mendeteksi investasi ilegal di Indonesia.

II. DASAR TEORI

Penipuan investasi di platform media sosial seperti Telegram telah menjadi masalah yang semakin meningkat dalam beberapa tahun terakhir. Putri dan Musyafak [9], meneliti jenis-jenis modus yang sering digunakan oleh pelaku penipuan di platform media sosial, termasuk Telegram. Mereka menemukan bahwa undian berhadiah dan penawaran investasi adalah modus yang paling sering digunakan. Studi ini menunjukkan bahwa pelaku penipuan sering menggunakan daya tarik keuntungan besar untuk menarik korban. Selain itu, pelaku penipuan juga sering menggunakan testimoni palsu atau memanipulasi data untuk menunjukkan bahwa investasi tersebut telah memberikan keuntungan besar kepada investor lain. Studi tersebut dapat dijadikan rujukan untuk mengidentifikasi pola penipuan investasi ilegal di Telegram.

Salah satu pendekatan yang dapat digunakan untuk mendeteksi penipuan investasi di Telegram adalah dengan menggunakan *text classification* yang merupakan salah satu metode dalam *machine learning*. Metode *text classification* memungkinkan komputer untuk mengategorikan teks ke dalam dua kategori, seperti ilegal atau legal.

Beberapa penelitian telah mengimplementasikan *machine learning* dalam *text classification*. Sebagai contoh, penelitian oleh Mutawalli dkk. [10] menggunakan metode (SVM) untuk melakukan *text classification* sosial media Twitter. Hasil penelitian menunjukkan bahwa SVM dapat menghasilkan akurasi yang cukup tinggi sebesar 80% dalam *text classification*.

Selain algoritma SVM, ada juga penelitian yang menggunakan algoritma lain untuk *text classification*. Contohnya, penelitian oleh Ardiani dkk [11] mengimplementasikan algoritma Naive Bayes untuk klasifikasi cuitan Twitter masyarakat terhadap pembangunan Kota Pontianak. Hasilnya menunjukkan bahwa Naive Bayes

mencapai tingkat akurasi sebesar 72% dari jumlah data latih sebesar 450 dan data uji sebesar 50.

Meskipun terdapat berbagai metode yang dapat digunakan dalam *text classification*, penggunaan algoritma LSTM memiliki performa yang lebih tinggi. Dalam sebuah penelitian yang dilakukan oleh Pratama [7], algoritma LSTM diimplementasikan untuk mengklasifikasikan teks data SMS spam berbahasa Indonesia. Hasil penelitian menunjukkan bahwa LSTM memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan dua algoritma lainnya, yaitu Naïve Bayes dan KNN. Algoritma LSTM mencapai tingkat akurasi sebesar 94%, sedangkan Naïve Bayes hanya sebesar 67% dan KNN sebesar 70%.

Studi serupa yang dilakukan oleh Nugroho dkk. [8] membandingkan kemampuan algoritma untuk mengidentifikasi kecemasan dan depresi pengguna Twitter. LSTM dan BiLSTM dibandingkan dengan semua model pembelajaran mesin tradisional. Pengujian dengan *confusion matrix* menunjukkan bahwa model algoritma tradisional seperti SVM, Naive Bayes, MLP, KNN, dan DT memiliki akurasi yang lebih rendah daripada LSTM dan BiLSTM. Namun, hasil pengujian yang dibandingkan menunjukkan bahwa BiLSTM memiliki akurasi 94.12%, sedangkan LSTM memiliki akurasi 84.91%.

Penelitian lain yang dilakukan oleh Fadli & Hidayatullah [12] mengimplementasikan BiLSTM dalam *text classification* untuk mengidentifikasi *cyberbullying* pada cuitan berbahasa Indonesia. Hasilnya menunjukkan bahwa BiLSTM mencapai tingkat akurasi 93.84% sedangkan dengan metode lain seperti LSTM tingkat akurasinya sebesar 92.02%. dan urutan kata yang kompleks dalam bahasa Indonesia.

Berdasarkan uraian di atas, sebagian besar penelitian terdahulu pada *text classification* menggunakan algoritma *machine learning*, seperti SVM, Naïve Bayes, dan LSTM. Sebagian besar penelitian tersebut masih berfokus pada *dataset* dari platform seperti Twitter dan SMS. Penelitian yang menggunakan *dataset* dari Telegram masih terbatas. Selain itu, algoritma BiLSTM dan LSTM memiliki performa yang lebih tinggi dibandingkan algoritma lainnya. Oleh karena itu, penggunaan algoritma LSTM dan Bi-LSTM dalam *text classification* perlu diuji performanya untuk mendeteksi penipuan investasi dari data pesan Telegram.

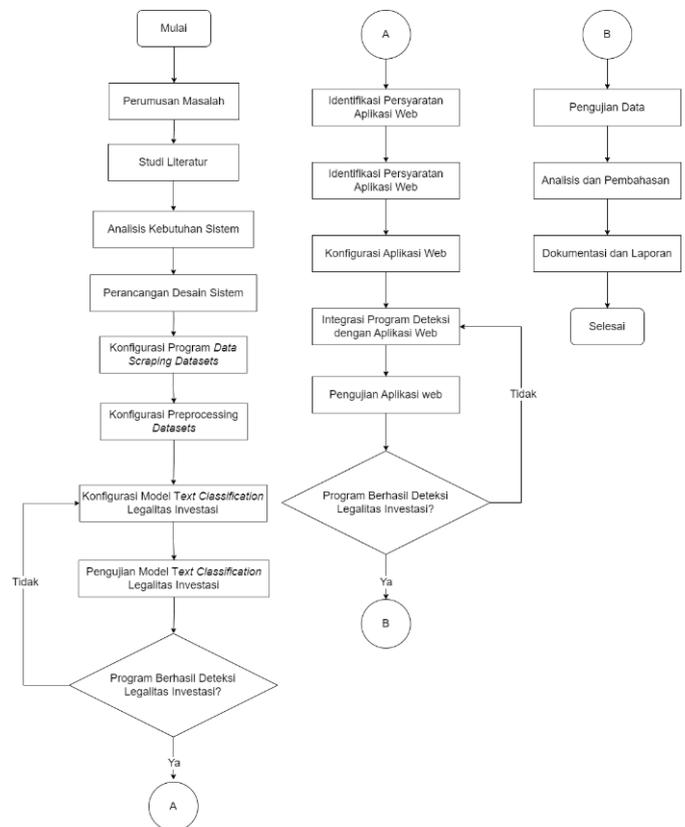
III. METODOLOGI

Penelitian ini terdiri dari beberapa tahapan yang dapat dilihat pada Gambar 1. Tahap awal adalah perumusan masalah dan studi literatur untuk memperdalam wawasan mengenai topik penelitian. Selanjutnya, kebutuhan sistem baik perangkat lunak maupun perangkat keras ditentukan. Tahap berikutnya adalah perancangan sistem dan instalasi program *scraper* untuk mengumpulkan *dataset* pesan investasi di Telegram. Setelah mengumpulkan *dataset*, langkah selanjutnya adalah melatih model *text classification* menggunakan algoritma BiLSTM dan LSTM, serta melakukan pengujian performa menggunakan metode

confusion matrix. Model dengan performa terbaik akan digunakan untuk purwarupa sistem deteksi investasi ilegal.

Selanjutnya, dilakukan pengujian *text classification* secara lokal menggunakan Flask API. Jika berhasil, dilanjutkan dengan pembuatan web interaktif menggunakan *framework* Ruby on Rails. Jika sistem deteksi tidak berhasil secara lokal, dilakukan *troubleshooting* pada konfigurasi program. Setelah berhasil membuat aplikasi web interaktif, dilakukan integrasi program deteksi dengan aplikasi web. Setelah integrasi selesai, dilakukan pengujian sistem deteksi menggunakan data yang telah disiapkan.

Sistem deteksi legalitas investasi diuji menggunakan data pesan iklan investasi dari *dataset* yang telah disiapkan. Hasil deteksi akan divisualisasikan pada halaman *website*, dan data hasil deteksi juga dapat diunduh dalam format CSV. Langkah terakhir adalah pembuatan laporan dan dokumentasi berdasarkan analisis dan pembahasan hasil pengujian sistem.



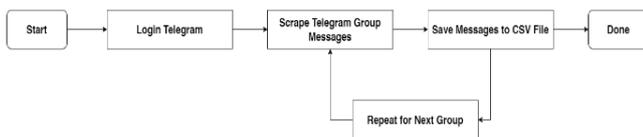
Gambar 1. Diagram alir

A. Alat dan Bahan

Dalam melaksanakan penelitian ini, dibutuhkan perangkat keras dan perangkat lunak guna menunjang penelitian yang akan dilakukan. Perangkat keras terdiri dari Laptop/PC dan SSD eksternal. Sedangkan perangkat lunak meliputi *Visual Studio Code*, Python, Google Collaboratory, PostgreSQL, Github, Tensorflow, Keras, Flask, Ruby on Rails.

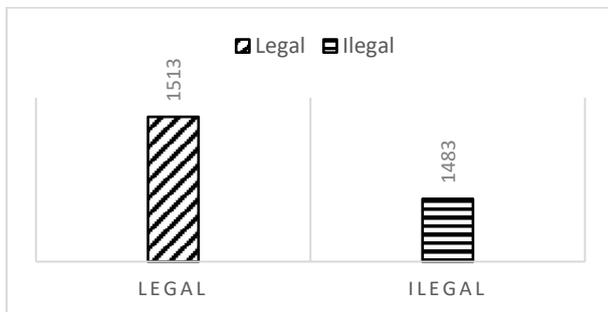
B. Pengumpulan Data

Tahap pengumpulan *dataset* dimulai dengan melakukan *data scraping* pesan iklan investasi dari grup Telegram. Diagram alir pada Gambar 2, menunjukkan langkah-langkah dalam program tersebut, di mana program *login* ke Telegram menggunakan API *credentials* yang valid. Selanjutnya, program mengambil pesan-pesan dari grup yang telah ditentukan sebelumnya. Pesan-pesan yang berhasil diambil kemudian disimpan dalam *file* CSV. Proses pengambilan pesan dari grup Telegram diulang hingga semua grup telah diproses. Setelah semua grup diproses, program akan berakhir.



Gambar 2. Diagram alir telegram *scraping*

Dua *dataset* legal dan ilegal hasil *scraping* pesan investasi dari grup Telegram yang masih terpisah diberikan label secara manual untuk menandakan pesan tersebut termasuk kategori legal atau ilegal. Kemudian, *dataset* yang telah diberikan label digabungkan menjadi sebuah *dataset* yang di dalamnya berisi pesan legal dan ilegal. *Dataset* tersebut terdapat dua jenis yang digunakan dalam pengujian model *text classification*, yaitu *dataset* yang berjumlah 2996 data (Gambar 3) dan 210 data (Gambar 4).



Gambar 3. Komposisi jumlah label legalitas *dataset* 2996 data



Gambar 4. Komposisi jumlah label legalitas *dataset* 210 data

C. Preprocessing

Dataset yang berhasil dikumpulkan akan melalui tahap *preprocessing* sebagai proses membersihkan *noise* dan beberapa hal yang tidak diperlukan dalam *dataset*. Langkah-langkah *preprocessing* yang digunakan meliputi [7] (i) *Folding case* (ii) *Filtering* (iii) *Tokenisasi* (iv) *Stemming* berbahasa Indonesia (v) *Stopwords Removal*.

Berdasarkan Tabel 1, dapat diidentifikasi bahwa langkah-langkah *preprocessing* yang telah dilakukan berhasil menghilangkan *noise* dan meningkatkan kualitas data untuk digunakan dalam model *text classification*.

Tabel 1. *Dataset* setelah melalui *preprocessing*

Teks	Legalitas
belum sempat buka aplikasi akseleran suka tinggal campaign jangan khawatir akseleran punya fitur utk kamu yang sangat sibuk tp tetap bisa laku dana nama auto lending apa sih auto lending	1
real testimoni terima kasih telah joint di channel kami moga dana manfaat ya kak rizki arisha yang mau ikut investasi dana silah japri admin juwai musah	0

D. Text classification

Data berupa pesan investasi diklasifikasi ke dalam kelas-kelas legalitas berupa kelas legal dan kelas ilegal menggunakan algoritma BiLSTM dan LSTM. Secara detail, dilakukan perbandingan performa dengan dua varian *dataset* yaitu, dengan jumlah datanya 2996 dan 210. Selain itu, jumlah *epoch* yang digunakan untuk setiap model divariasikan dengan kelipatan 8 dengan nilai *epoch* paling rendah 16 dan paling tinggi 72. Hal tersebut dikarenakan nilai *epoch* yang paling sesuai untuk model tidak dapat dipastikan sehingga perlu menguji beberapa *epoch* untuk mencapai akurasi yang optimal. Proses pelatihan model dilakukan berulang-ulang untuk mencapai konvergensi nilai bobot. Rasio pembagian *data training* dan *data testing* yang digunakan adalah 30:70, 50:50, dan 70:30.

E. Evaluasi

Confusion matrix adalah sebuah matriks yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi dan data aktualnya. Untuk menghitung kinerja model, dapat digunakan metrik seperti akurasi, presisi, dan *recall* berdasarkan rumus yang tercantum pada persamaan (1), (2), dan (3). Rumus-rumus tersebut adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

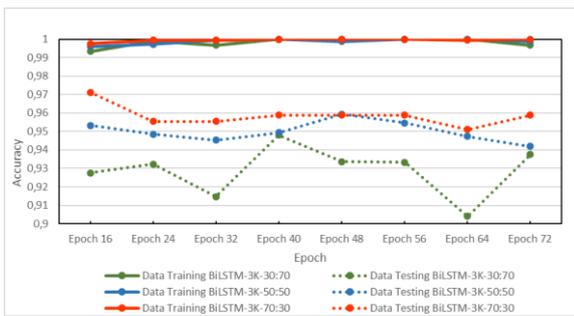
$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Model

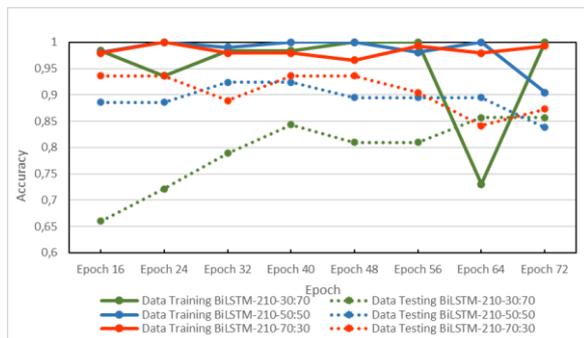
Pengujian pelatihan model dilakukan untuk mengetahui pengaruh jumlah *epoch* dan rasio pembagian data terhadap performa model dalam mengklasifikasikan pesan investasi selama pelatihan. Variasi jumlah *epoch* adalah 8, 16, 24, 32, 40, 48, 56, 64, dan 72. Sedangkan rasio pembagian *data training* dan *data testing* yang digunakan adalah 30:70, 50:50, dan 70:30.

Hasil pengujian pada varian model BiLSTM dengan 2996 data, seperti yang terlihat dalam Gambar 5, menunjukkan tingkat akurasi yang cenderung tinggi pada *epoch* 16, 48, dan 72. Namun, perlu diperhatikan bahwa pada *epoch* 72, varian dengan rasio pembagian 50:50 lebih akurat daripada varian lainnya. Pada *epoch* tersebut, model telah mencapai atau mendekati konvergensi dimana penyesuaian bobot model telah mencapai puncaknya. Selain itu, tingkat akurasi yang dicapai meningkat seiring dengan jumlah data latihan yang digunakan. Dengan rasio pembagian 70:30 antara data latihan dan pengujian, tingkat akurasi yang dicapai lebih tinggi.



Gambar 5. Perbandingan performa latih model BiLSTM varian 2996 data

Jika dibandingkan dengan model varian BiLSTM yang memiliki jumlah data 2996. Hasil pengujian model varian BiLSTM dengan jumlah data 210 berbeda. Hasil tersebut ditampilkan dalam Gambar 6.



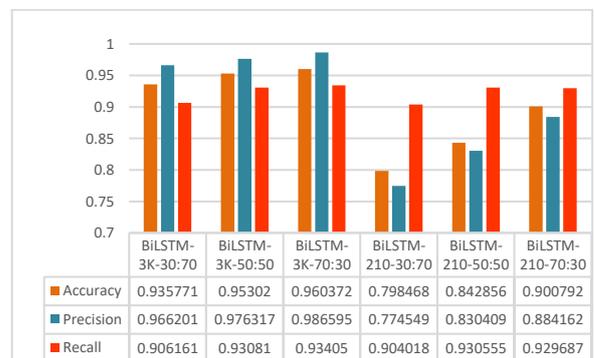
Gambar 6. Perbandingan performa latih model BiLSTM varian 210 data

Hasil pengujian model varian BiLSTM dengan jumlah data 210 digambarkan pada Gambar 6. Dalam pengujian varian model tersebut, model menunjukkan tingkat akurasi

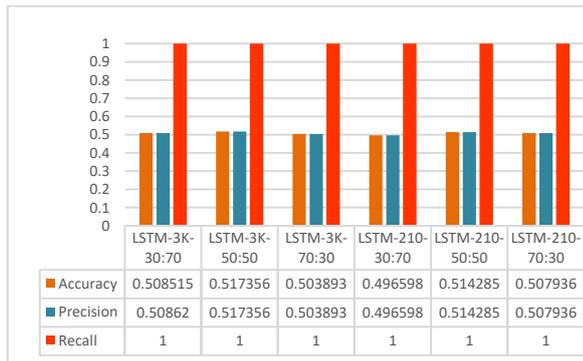
yang tinggi pada *epoch* 40, tetapi kemudian menurun setelahnya. Akurasi meningkat sedikit pada *epoch* 72 yang menunjukkan bahwa model telah mencapai atau hampir mencapai titik konvergensi saat pelatihan dilakukan pada *epoch* 40. Temuan juga menunjukkan bahwa semakin banyak data pelatihan yang digunakan dalam model, maka hasilnya semakin akurat. Hal ini ditunjukkan dengan perbandingan data 70:30 antara data pelatihan dan pengujian. Namun, model dengan jumlah 210 data ini menunjukkan kinerja yang lebih buruk dengan akurasi hanya 0,843. Hasil akurasi tersebut didapatkan dari menghitung rata-rata semua akurasi model dengan jumlah 210 data.

Pengujian selanjutnya adalah untuk mengetahui performa akurasi, presisi, dan *recall data testing* dengan pengukuran *confusion matrix*. Pengujian *confusion matrix* pada model *text classification* dilakukan pada dua algoritma, LSTM dan BiLSTM. Masing-masing algoritma diuji dengan dua varian *dataset*, masing-masing dengan 2996 data dan 210 data. Selain itu, *dataset* ini juga diuji dengan berbagai rasio pembagian data pelatihan dan pengujian, yaitu 30:70, 50:50, dan 70:30.

Berdasarkan Gambar 8, model LSTM tanpa *layer bidirectional* tidak mampu mengekstraksi informasi konteks secara optimal sehingga performanya rendah dibandingkan dengan model BiLSTM. Hal tersebut terlihat pada Gambar 8, bahwa semua performa varian model LSTM konstan sama. Hasil tersebut berbanding terbalik dengan performa varian model BiLSTM seperti tertampil pada Gambar 7. Varian model dengan split *dataset* rasio 70:30 dan *dataset* berjumlah 2996 data menghasilkan tingkat akurasi (96%), presisi (98%), dan *recall* (93,4%) yang lebih tinggi dibandingkan dengan varian lain dari model BiLSTM dan model LSTM. Hal ini mengindikasikan bahwa penggunaan *layer bidirectional*, jumlah *dataset* yang lebih banyak, dan rasio split *dataset* 70:30 memberikan performa signifikan yang berdampak pada peningkatan akurasi, presisi, dan *recall* dalam klasifikasi pesan investasi.



Gambar 7. Hasil perbandingan performa model BiLSTM

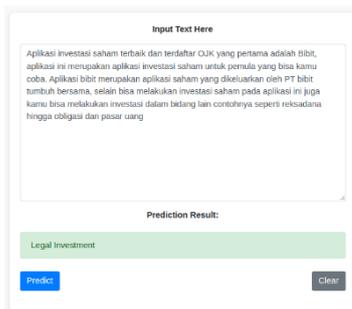


Gambar 8. Hasil perbandingan performa model LSTM

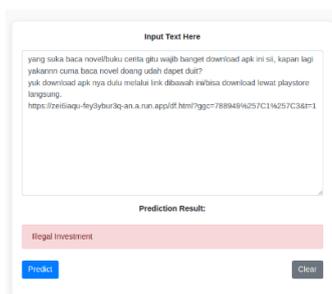
B. Hasil Pengujian Fungsionalitas Situs

Setelah menguji performa dari semua varian model, maka selanjutnya adalah menyimpan varian model dengan performa tertinggi ke dalam *file* berformat H5. Model tersebut digunakan sebagai API prediksi oleh sistem berbasis web.

Berdasarkan Gambar 9 dan 10, situs purwarupa ini berfungsi dengan baik. Situs dapat memprediksi pesan investasi ilegal dan legal. Selain itu, API model *text classification* dapat memberikan respons yang cepat dan efektif terhadap masukan teks pesan investasi.



Gambar 9. Hasil prediksi investasi legal



Gambar 10. Hasil prediksi investasi ilegal

V. SIMPULAN

Berdasarkan hasil pengujian yang telah dicapai, dapat disimpulkan bahwa Purwarupa Sistem Pencegahan Penipuan Investasi dengan Metode *Text classification* Berbasis Algoritma *Bidirectional LSTM* berhasil dirancang dan mampu memprediksi pesan investasi dengan performa akurasi, presisi, dan *recall* yang tinggi. Model yang menghasilkan performa tertinggi adalah varian model yang

menggunakan algoritma BiLSTM dengan jumlah 2996 data dan rasio pembagian *dataset* 70:30. Penelitian yang telah dilakukan masih memiliki beberapa kekurangan yang perlu diperbaiki dan dikembangkan di masa mendatang. Hal yang perlu dikembangkan adalah menambah *dataset*, menambah pengujian *epoch* dan *split dataset*, dan mengoptimalkan performa model yang menggunakan algoritma LSTM dengan menambahkan beberapa parameter.

REFERENSI

- [1] KSEI, "Pencapaian Tahun 2022 yang Positif, Tumbuhkan Semangat Baru pada Tahun 2023," 2022. [Online]. Available: https://www.ksei.co.id/files/uploads/press_releases/press_file/id-id/215_berita_pers_pencapaian_tahun_2022_yang_positif_tumbuhkan_semangat_baru_pada_tahun_2023_20230102182734.pdf.
- [2] W. Albrecht, C. A. C. Albrecht and M. Zimelman, *Fraud examination: Cengage Learning*, 2018.
- [3] cnnindonesia.com, "Rugi Masyarakat Akibat Investasi Bodong Melesat Jadi Rp109 T di 2022," 2022. [Online]. Available: <https://www.cnnindonesia.com/ekonomi/20221117193808-78-875287/rugi-masyarakat-akibat-investasi-bodong-melesat-jadi-rp109-t-di-2022>.
- [4] Databoks, "Media Sosial yang Digunakan Responden untuk Mencari Informasi Produk Investasi," Oktober 2022. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2022/12/01/ini-media-sosial-paling-efektif-untuk-promosi-produk-investasi>.
- [5] K. Kowsari, D. E. Brown, M. H., K. J. M., M. S. Gerber and L. E. Barnes, "HDLTex: Hierarchical Deep Learning for Text Classification," *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 364-371, 2017.
- [6] F. Rahutomo, I. Y. R. Pratiwi and D. M. Ramadhani, "Eksperimen Naïvebayes pada Deteksi Berita Hoax Berbahasa Indonesia," 2019. [Online]. Available: <https://jurnal.kominfo.go.id/index.php/jpkop/article/view/1805/1122>
- [7] E. D. Pratama, "Implementasi Model Long-Short Term Memory (LSTM) pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia. The Journal on Machine Learning and Computational Intelligence (JMLCI), 1(2)," 2022. Available: <https://jmlci.unesa.ac.id/index.php/home/article/view/12/>.
- [8] K. S. Nugroho, I. Akbar and I. Affi Nizar S., "Deteksi Depresi dan Kecemasan Pengguna Twitter Menggunakan Bidirectional LSTM," 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2301.04521>.
- [9] F. A. Putri and N. Musyafak, "Pesan Komunikasi dalam Penipuan Digital Pada Masa Pandemi COVID-19," 2022. [Online]. Available: https://www.researchgate.net/profile/Najahan-Musyafak/publication/368660294_Pesan_Komunikasi_Dalam_Penipuan_Digital_Pada_Masa_Pandemi_Covid-19_Communication_Messages_In_Digital_Deception_During_The_Covid-19_Pandemic/links/63f36c6731cb6a6d1d19b551/PESAN-KO.
- [10] L. Mutawalli, M. T. A. Zaen and W. Bagye, "Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)," 2019. [Online]. Available: <https://ejournal.stmiklombok.ac.id/index.php/jire/article/view/117/77>.
- [11] L. Ardiani, H. Sujaini and Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 8(2), pp. pp.183-190, 2020.
- [12] H. F. Fadli and A. F. Hidayatullah, "Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM," 2023. [Online]. Available: <https://journal.uui.ac.id/AUTOMATA/article/view/17364/10897>.