

Prediksi Gender Berdasarkan Nama Bahasa Indonesia Menggunakan *Long Short Term Memory*

<http://dx.doi.org/10.28932/jutisi.v9i2.6404>

Riwayat Artikel

Received: 6 Agustus 2023 | Final Revision: 19 Agustus 2023 | Accepted: 19 Agustus 2023

Creative Commons License 4.0 (CC BY – NC)



Arya Mulya Kusuma^{#1}, Harisatul Aulia^{#2}, Muhammad Alfaris Oktavian^{#3}, Muhammad Rizky Akbar^{#4}, Putri Patricia^{#5}, Abdiansah Abdiansah^{✉#6}

[#] Program Studi Sarjana Ilmu Komputer, Universitas Sriwijaya, Indralaya

Jl. Indralaya Utara, Kabupaten Ogan Ilir, 30862, Indonesia

¹aryakusuma832@gmail.com

²haris.aulia404@gmail.com

³alfarisoktavian99@gmail.com

⁴rizky7754@gmail.com

⁵ppatricia.syah@gmail.com

⁶abdiansah@unsri.ac.id

✉ Corresponding author: abdiansah@unsri.ac.id

Abstrak — Prediksi gender merupakan salah satu bentuk dari program yang dapat melakukan prediksi yang mengeluarkan output berupa jenis gender. Pada riset yang dilakukan pada studi kasus ini, digunakan input berupa teks nama orang Indonesia. Dalam penelitian ini, dilakukan eksperimen menggunakan metode *Long Short Term Memory* untuk melakukan prediksi dan klasifikasi terhadap nama-nama orang Indonesia untuk mengetahui gender mereka hanya berdasarkan nama dengan tujuan untuk meningkatkan ilmu pengetahuan dan mencari inovasi baru untuk menopang penelitian-penelitian yang akan datang. Batasan penelitian ini adalah penggunaan dataset yang terdiri dari nama-nama orang Indonesia dan pengklasifikasian jenis kelamin hanya terdiri dari dua kategori, yaitu laki-laki dan perempuan. Perbandingan akurasi dari hasil training program baseline dan program hasil modifikasi adalah akurasi untuk program baseline sekitar 0.93, sedangkan akurasi untuk program hasil modifikasi sekitar 0.90. Hasil penelitian menunjukkan adanya peningkatan akurasi setelah melakukan percobaan dengan data testing pada program hasil modifikasi yaitu sebesar 0.96.

Kata kunci — Gender; Long Short Term Memory; Nama Bahasa Indonesia; Prediksi.

Gender Prediction Based on Indonesian Names Using Long Short Term Memory

Abstract — Gender prediction is a form of program that can make predictions that issue output in the form of gender. In the research conducted in this case study, the input was in the form of Indonesian names. To improve knowledge and seek new innovations to support future research. The limitation of this research is the use of a dataset consisting of Indonesian names and the classification of gender only consists of two categories, namely male and female. Comparison of the accuracy of the results of the baseline program training and the modified program is that the accuracy for the baseline program is around 0.93, while the accuracy for the modified program is around 0.90. The results showed that there was an increase in accuracy after conducting experiments with data testing on the modified program, which was equal to 0.96.

Keywords — Gender; Indonesian Names; Long Short Term Memory; Prediction.

I. PENDAHULUAN

Prediksi jenis kelamin dari nama merupakan topik yang semakin menarik perhatian banyak ahli di bidang pemrosesan bahasa alami [1]. Terdapat berbagai metode yang dapat digunakan untuk memprediksi jenis kelamin berdasarkan nama, seperti metode *Long Short Term Memory* (LSTM) dan *Machine Learning* [2], [3]. Meskipun masalah ini terlihat sederhana, namun prediksi *gender* dari nama dapat menjadi sulit karena adanya ambiguitas dalam nama dan bahasa yang digunakan [4]. Sebagai solusinya, metode LSTM telah terbukti efektif dalam menangani masalah prediksi jenis kelamin dari nama dalam Bahasa Inggris [5], [6] dan bahasa lainnya [7], [8], [9], [10], [11]. Akan tetapi, masih terbatas penelitian yang memprediksi *gender* dari nama dalam Bahasa Indonesia menggunakan metode LSTM.

Penelitian ini akan menggunakan metode LSTM untuk mempelajari pola *sekuensial* dalam nama dan karakteristik *gender* [8], [9], [12]. Selain itu, juga akan diperkenalkan teknik baru untuk *encoding character-level* dan *word-level* dari nama, menggunakan teknik *embedding* [13]. Tujuan dari penelitian ini adalah untuk menghasilkan model prediksi *gender* yang akurat dan dapat dipercaya berdasarkan nama Bahasa Indonesia.

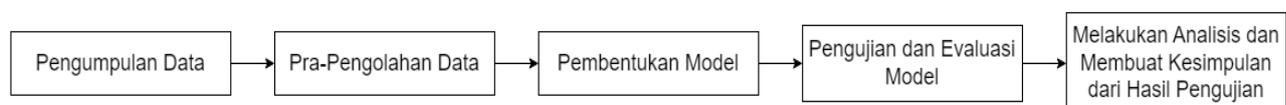
Sebuah studi [2] digunakan sebagai *baseline* dalam penelitian ini, yang mengembangkan model untuk memprediksi jenis kelamin dari nama Bahasa Indonesia menggunakan metode *Naive Bayes*. Hasil studi menunjukkan bahwa model yang dikembangkan dapat mencapai tingkat akurasi yang cukup baik, yaitu sekitar 93%. Namun, penelitian-penelitian sebelumnya telah menunjukkan bahwa metode-metode tradisional seperti SVM [14], *Naive Bayes* [2], [14], dan *Decision Tree* [15], [16] memiliki keterbatasan dalam menangani ketergantungan jangka panjang dalam data *sekuensial*, sedangkan LSTM adalah jenis jaringan saraf *rekursif* yang dirancang untuk mengatasi masalah tersebut [4], [17], [18].

Beberapa penelitian terbaru telah menggunakan metode LSTM untuk memprediksi *gender* berdasarkan nama dalam berbagai bahasa seperti Bahasa China [7], Prancis [8], Inggris [9], Brazil [10], dan India [11]. Namun, masih sedikit penelitian yang telah dilakukan pada prediksi *gender* berdasarkan nama Bahasa Indonesia menggunakan LSTM [2], [16]. Oleh karena itu, penelitian ini akan memberikan kontribusi baru dalam mengembangkan metode prediksi *gender* yang lebih baik dan dapat diandalkan untuk Bahasa Indonesia menggunakan LSTM. Dengan demikian, penelitian ini diharapkan dapat memberikan manfaat bagi pengembangan aplikasi dan sistem berbasis Bahasa Indonesia yang lebih efektif dan efisien dalam mengidentifikasi jenis kelamin berdasarkan nama.

Tujuan utama dari penelitian ini adalah mengembangkan sebuah model prediksi jenis kelamin berdasarkan nama dalam Bahasa Indonesia dengan menggunakan metode *Long Short Term Memory*. Penelitian ini bertujuan untuk menginvestigasi pola *sekuensial* yang terdapat dalam nama serta karakteristik yang terkait dengan *gender*, sekaligus memperkenalkan teknik baru untuk melakukan *encoding* pada tingkat karakter dan kata menggunakan teknik *embedding*. Fokus utama penelitian ini adalah menghasilkan sebuah model prediksi *gender* yang memiliki akurasi tinggi dan dapat diandalkan ketika diterapkan pada nama-nama dalam Bahasa Indonesia. Selain itu, penelitian ini juga bertujuan untuk mengatasi keterbatasan yang dimiliki oleh metode-metode tradisional dalam menghadapi ketergantungan jangka panjang dalam data *sekuensial*, dengan memanfaatkan keunggulan LSTM sebagai jaringan saraf *rekursif* yang dirancang khusus untuk menangani masalah tersebut. Melalui kontribusinya dalam mengembangkan metode prediksi *gender* yang lebih unggul dan dapat diandalkan untuk nama-nama dalam Bahasa Indonesia, diharapkan bahwa penelitian ini dapat memberikan dukungan yang signifikan bagi pengembangan aplikasi dan sistem berbasis Bahasa Indonesia yang efektif dan efisien dalam mengidentifikasi jenis kelamin berdasarkan nama.

II. METODE PENELITIAN

A. Tahapan Penelitian



Gambar 1. Alur Tahapan Penelitian

Penelitian ini melibatkan beberapa tahapan penting yang tertuang pada gambar 1. Tahap pertama adalah pengumpulan data, di mana informasi relevan diperoleh dari berbagai sumber terkait. Kemudian, data tersebut menjalani pra-pengolahan, dimana proses pembersihan dan organisasi dilakukan untuk mempersiapkan data yang akan digunakan. Setelah itu, dilakukan pembentukan model dengan menggunakan data yang telah diproses. Tahap selanjutnya adalah pengujian dan evaluasi model, di mana model tersebut diuji dengan skenario *testing* menggunakan 100 data sampel yang belum pernah digunakan sebelumnya untuk mengukur kinerja dan keandalannya. Hasil pengujian analisis dan kesimpulan diambil untuk menyusun temuan dari penelitian tersebut. Dengan alur ini, penelitian dapat dilakukan secara sistematis dan menghasilkan hasil yang dapat diandalkan dan relevan dengan tujuan penelitian yang telah ditetapkan.

B. Dataset

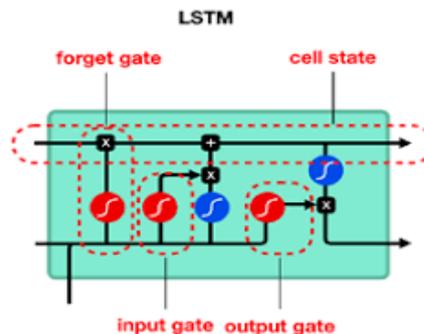
Untuk membangun model program prediksi *gender* berdasarkan nama Bahasa Indonesia dengan metode *Long Short Term Memory*, digunakan *dataset* sekunder. *Dataset* tersebut diperoleh dari Komisi Pemilihan Umum (KPU) dan berisi informasi Data Pemilih pada Pilkada DKI Jakarta tahun 2022, termasuk nama dan jenis kelamin individu. Data tersebut digunakan sebagai data pelatihan dalam pengembangan model program.

Selanjutnya, untuk melakukan pengujian program, digunakan *dataset* sekunder lain yang berasal dari Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya (Fasilkom Unsri). *Dataset* ini terdiri atas informasi nama dan jenis kelamin mahasiswa Teknik Informatika Fasilkom Unsri. Dengan menggunakan kedua jenis *dataset* tersebut, model program dapat diuji keakuratannya dalam memprediksi jenis kelamin berdasarkan nama Bahasa Indonesia.

Sebelum melakukan pengujian, dilakukan proses pelatihan pada model program menggunakan data pelatihan dari *dataset* Pilkada DKI Jakarta tahun 2022. Melalui proses pelatihan tersebut, konfigurasi parameter yang optimal dapat ditentukan agar model program memberikan hasil prediksi yang akurat. Setelah itu, dilakukan pengujian program dengan menggunakan *data testing* dari *dataset* nama mahasiswa Teknik Informatika Fasilkom Unsri dan hasilnya dibandingkan dengan program *baseline* yang menggunakan metode *Naive Bayes Classifier*.

C. Long Short Term Memory

Metode yang digunakan yaitu *Long Short Term Memory* (LSTM). LSTM adalah salah satu jenis dari *Recurrent Neural Network* (RNN) dimana dilakukan modifikasi pada RNN dengan menambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama [19]. RNN rentan terhadap *vanishing* dan *exploding gradients* sehingga LSTM diperkenalkan dan menjadi salah satu varian paling populer dari RNN. Permasalahan *vanishing gradient* ini mengakibatkan RNN gagal dalam menangkap *long term dependencies* [20], sehingga mengurangi akurasi dari suatu prediksi pada RNN [21].



Gambar 2. Arsitektur *Long Short Term Memory* [22]

Pada Gambar 2 [22], ditampilkan ilustrasi atau representasi visual arsitektur LSTM yang digunakan dalam proses analisis dan pemodelan. LSTM mengatasi permasalahan *vanishing gradient* pada RNN dengan menggunakan tiga gerbang yaitu *forget gate*, *input gate*, dan *output gate* [23]. *Forget gate* berfungsi untuk memilih informasi apa yang perlu dilupakan dari *memory cell*, sedangkan *input gate* berfungsi untuk menentukan informasi baru apa yang perlu disimpan dalam *memory cell* [24]. *Output gate* berfungsi untuk memilih informasi mana yang akan digunakan sebagai *output* dari LSTM [25]. Pada langkah awal, *forget gate* memutuskan apakah informasi akan dipertahankan atau dilupakan dengan menggunakan fungsi *sigmoid*. Nilai antara 0 hingga 1 menentukan apakah informasi akan diteruskan atau dilupakan. Kemudian, *input gate* mengupdate nilai *cell state* dengan menggunakan informasi dari *input* sebelumnya (X) dan *gate input* (I) melalui fungsi *sigmoid* dan *tanh*. *Sigmoid* memutuskan informasi yang perlu di *update*, sementara *tanh* merepresentasikan perubahan nilai pada *cell state*. Perubahan nilai *cell state* dilakukan dengan penjumlahan informasi yang melewati *cell state*. Setelah itu, nilai *cell state* dihitung dengan mengalikan *output forget gate* dan menjumlahkan *output input gate*. Gerbang ini penting untuk menentukan *input* pada *state* selanjutnya dengan menggunakan *input* dari sebelumnya melalui fungsi *sigmoid*. Hasilnya dikalikan dengan perkalian antara *cell state* baru dan fungsi aktivasi *tanh* [26]. Dengan adanya *memory cell* dan tiga gerbang ini, LSTM dapat mengatasi permasalahan RNN dalam menangkap *long term dependencies*. Sehingga, LSTM dapat menghasilkan prediksi yang lebih akurat dalam tugas-tugas yang melibatkan urutan data seperti *speech recognition*, *natural language processing*, dan sebagainya [27].

LSTM dapat digunakan untuk tugas prediksi nama, seperti memprediksi jenis kelamin atau kebangsaan dari nama yang diberikan [28]. Dalam tugas-tugas ini, model LSTM mengambil karakter dari nama sebagai masukan dan belajar memprediksi label yang sesuai (misalnya, laki-laki atau perempuan untuk prediksi jenis kelamin).

Untuk membangun model LSTM untuk prediksi nama, data *input* perlu diproses terlebih dahulu dan dikodekan sebagai rangkaian karakter atau *token*. Model LSTM kemudian dapat dilatih pada data yang disandikan menggunakan pendekatan pembelajaran terawasi, di mana urutan masukan dipasangkan dengan label yang sesuai.

Setelah model LSTM dilatih, model ini dapat digunakan untuk membuat prediksi pada nama baru dengan memasukkan karakter nama sebagai *input* dan mendapatkan label prediksi sebagai *output* [29]. Keakuratan prediksi bergantung pada kualitas dan ukuran data pelatihan, serta kompleksitas dan arsitektur model LSTM. Secara keseluruhan, LSTM dapat menjadi alat yang berguna untuk tugas prediksi nama, terutama saat menangani kumpulan data besar dan pola kompleks dalam data masukan.

D. Percobaan Baseline

Pada percobaan *baseline* (rujukan), metode yang digunakan dalam melakukan prediksi *gender* yaitu metode *Naive Bayes Classifier*. *Naive Bayes* adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi [30]. Ini adalah algoritma *probabilistik* yang memodelkan probabilitas setiap kelas yang diberikan serangkaian fitur masukan [31].

Naive Bayes dan LSTM adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi, tetapi keduanya berbeda dalam pendekatan dan kinerjanya [32]. Dalam hal kinerja, hasilnya bervariasi tergantung pada tugas dan kumpulan data tertentu. Beberapa penelitian menemukan bahwa LSTM mengungguli *Naive Bayes* dalam tugas klasifikasi tertentu, seperti deteksi berita palsu dan analisis sentimen [33]. Namun, penelitian lain menemukan bahwa *Naive Bayes* dapat bekerja sebanding atau bahkan lebih baik daripada LSTM dalam beberapa kasus, seperti mengklasifikasikan tweet pelecehan seksual [8].

Pada metode *baselinenya*, digunakan *classifier multinomial naive bayes* dan digunakan juga ekstraksi fitur *count vectorizer* dengan *analyzer char_wb*. Tingkat akurasi yang didapatkan pada percobaan *baseline* menghasilkan akurasi sebesar 93.3%.

III. HASIL DAN PEMBAHASAN

A. Percobaan Modifikasi

Dalam penelitian ini, dilakukan perubahan pada metode yang digunakan, dari *Naive Bayes classifier* sebagai metode dasarnya. Perubahan tersebut dilakukan dengan menerapkan metode LSTM pada sumber kode *baseline*. Selain itu, *dataset* yang digunakan merupakan *dataset* yang sama seperti yang digunakan dalam sumber kode *baseline*, yaitu *dataset* yang disediakan oleh Komisi Pemilihan Umum (KPU). Dengan demikian, penelitian ini memodifikasi metode yang digunakan namun tetap menggunakan *dataset* yang sama dengan penelitian sebelumnya.

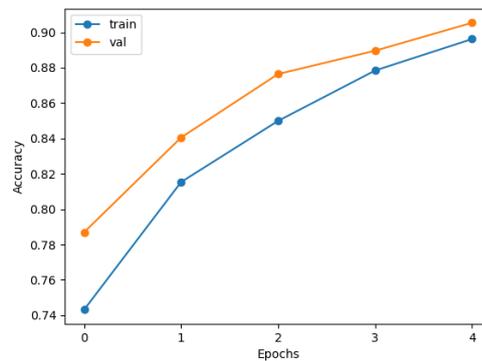
TABEL 1
KONFIGURASI DAN HASIL

| Konfigurasi | Hasil |
|-------------------|----------------|
| <i>Batch Size</i> | 64 |
| <i>Epoch</i> | 5 |
| <i>Embed Dim</i> | 256 |
| <i>Units</i> | 128 |
| <i>Drop out</i> | 0.2 |
| <i>Activation</i> | <i>sigmoid</i> |
| <i>Optimizer</i> | <i>Adam</i> |
| <i>Dense</i> | 1 |

Dalam penggunaan model LSTM pada penelitian ini, penyesuaian dilakukan pada beberapa parameter yang diterapkan agar menghasilkan akurasi yang lebih baik. Untuk memberikan gambaran lebih jelas mengenai konfigurasi parameter yang diterapkan, daftar parameter tersebut telah disusun pada Tabel 1 yang dapat dilihat pada laporan penelitian. Dengan melakukan penyesuaian parameter pada model LSTM yang digunakan, diharapkan hasil yang dihasilkan dapat lebih optimal dan akurat dalam memprediksi variabel yang menjadi fokus penelitian.

Setelah melakukan penyesuaian parameter pada model LSTM yang digunakan, selanjutnya melakukan proses pelatihan (*training*) menggunakan *dataset* yang sama dengan konfigurasi parameter yang sudah disesuaikan sebelumnya. Berdasarkan hasil yang diperoleh, tercapai tingkat akurasi yang cukup tinggi, yaitu sekitar 0.90, dengan nilai *loss* sebesar 0.24. Hal ini menunjukkan bahwa penyesuaian parameter yang dilakukan pada model LSTM berdampak positif terhadap

akurasi hasil prediksi yang dihasilkan. Akan tetapi hasil penyesuaian yang telah dilakukan belum berhasil melampaui dari kode *baseline*(rujukan) dimana kode rujukan memiliki akurasi sebesar 0.03 lebih besar dari akurasi kode modifikasi.



Gambar 3. Grafik Akurasi

Gambar 3 merupakan grafik yang menampilkan akurasi hasil analisis dari proses pelatihan model. Pada grafik tersebut, diperoleh akurasi awal sebesar 0.7431 dan validasi akurasi awal sebesar 0.7869. Namun, terlihat pada grafik bahwa seiring dengan peningkatan iterasi pelatihan (*epoch*), terjadi kenaikan yang signifikan dalam performa model. Akurasi secara bertahap meningkat menjadi 0.8153, 0.8499, 0.8784, dan akhirnya mencapai 0.9054. Begitu pula dengan validasi akurasi yang meningkat menjadi 0.8405, 0.8764, dan mencapai puncaknya pada 0.9054. Grafik tersebut menggambarkan perubahan positif yang menunjukkan peningkatan akurasi pada setiap iterasi pelatihan. Dengan pola peningkatan ini, dapat disimpulkan bahwa model mampu mengenali pola-pola yang semakin kompleks seiring berjalannya waktu. Hasil ini menggambarkan seiring banyak iterasi pelatihan yang dilakukan, semakin baik pula kemampuan model dalam memprediksi variabel.

B. Hasil Percobaan Baseline

Dari kode *baseline* yang menggunakan metode *Naive Bayes Classifier*, telah dilakukan beberapa percobaan menggunakan data *testing* dengan hasil percobaan sebagai berikut:

TABEL 2
HASIL PERCOBAAN BASELINE

| No. | Nama | Gender Asli | Hasil Prediksi |
|-----|------------------------|-------------|----------------|
| 1 | Achmad Mario Sunjabar | Laki-Laki | Laki-Laki |
| 2 | Adinda Septiana | Perempuan | Perempuan |
| 3 | Afifah Putri Mevtilina | Perempuan | Perempuan |
| 4 | Agung Dwi Paradisafa | Laki-Laki | Laki-Laki |
| 5 | Agustina Rosaria Saud | Perempuan | Perempuan |
| ... | | | |
| 100 | Affandi Arrizal | Laki-Laki | Laki-Laki |

Dari hasil pengujian program *baseline* yang menggunakan metode *Naive Bayes* pada 100 data uji, diperoleh informasi bahwa program tersebut berhasil memprediksi dengan benar 93 data uji dari total 100 data uji yang tersedia. Meskipun program ini berhasil memberikan tingkat akurasi yang cukup tinggi, namun terdapat 7 data uji yang tidak berhasil diprediksi dengan benar oleh program tersebut. Hal ini menunjukkan bahwa program *baseline* dengan metode *Naive Bayes* masih memiliki kelemahan dalam memprediksi beberapa kasus, terutama dalam memprediksi data yang memiliki karakteristik yang lebih kompleks atau sulit diprediksi.

C. Hasil Percobaan Modifikasi

Setelah menerapkan metode LSTM sebagai modifikasi pada penelitian ini, dilakukan serangkaian percobaan untuk mengevaluasi kualitas dan performa model yang dihasilkan. Berdasarkan hasil percobaan yang dilakukan, didapatkan hasil yang cukup memuaskan dalam memprediksi nama. Metode LSTM yang diterapkan dalam penelitian ini mampu memprediksi nama-nama pada data *testing* dengan cukup baik. Hasil percobaan tersebut dapat dilihat pada Tabel 3.

TABEL 3
HASIL PERCOBAAN MODIFIKASI

| No. | Nama | Gender Asli | Hasil Prediksi | Akurasi |
|-----|------------------------|-------------|----------------|---------|
| 1 | Achmad Mario Sunjabar | Laki-Laki | Laki-Laki | 1.00 |
| 2 | Adinda Septiana | Perempuan | Laki-Laki | 0.51 |
| 3 | Afifah Putri Mevtilina | Perempuan | Perempuan | 0.99 |
| 4 | Agung Dwi Paradisafa | Laki-Laki | Laki-Laki | 0.99 |
| 5 | Agustina Rosaria Saud | Perempuan | Perempuan | 0.91 |
| ... | | | | |
| 100 | Affandi Arrizal | Laki-Laki | Perempuan | 1.00 |

Skenario *testing* dilakukan evaluasi terhadap model prediksi yang telah dimodifikasi dengan menggunakan metode LSTM pada data *testing* yang terdiri dari 100 sampel. Hasil pengujian menunjukkan bahwa program berhasil memprediksi dengan benar 96 data uji dari total 100 data uji yang tersedia, mencapai tingkat akurasi sebesar 96%. Hal ini menunjukkan bahwa penggunaan metode LSTM pada program berhasil meningkatkan tingkat akurasi program dalam memprediksi nama jika dibandingkan dengan penggunaan metode *Naive Bayes* pada kode *baseline*.

Meskipun pada saat *training*, akurasi kode modifikasi yang menggunakan metode LSTM lebih kecil dibandingkan dengan kode *baseline* yang menggunakan metode *Naive Bayes*, namun hasil pengujian pada data *testing* menunjukkan hasil yang lebih baik pada kode modifikasi dengan metode LSTM. Oleh karena itu, penggunaan metode LSTM pada program dapat menjadi alternatif yang lebih efektif dan akurat dalam memprediksi variabel yang kompleks atau sulit diprediksi. Namun, perlu diingat bahwa penggunaan metode yang tepat dan konfigurasi parameter yang optimal sangat penting dalam menghasilkan model yang akurat dan dapat diandalkan dalam memprediksi nama.

IV. SIMPULAN

Perbandingan antara program *baseline* dan program hasil modifikasi, diperoleh informasi bahwa akurasi untuk program *baseline* mencapai sekitar 0.93, sedangkan akurasi untuk program hasil modifikasi sekitar 0.90. Namun, setelah dilakukan percobaan dengan menggunakan data *testing* pada kedua program tersebut, ditemukan bahwa hasil yang diperoleh menunjukkan bahwa tingkat keakuratan program hasil modifikasi dengan metode *Long Short Term Memory* (LSTM) lebih baik. Dalam menguji 100 nama yang berasal dari data Mahasiswa Teknik Informatika Fasilkom UNSRI 2020, program hasil modifikasi berhasil memprediksi dengan benar 96 nama, sedangkan program *baseline* hanya berhasil memprediksi 93 nama. Terdapat 4 nama yang diprediksi salah pada program hasil modifikasi, sedangkan pada program *baseline* terdapat 7 nama yang diprediksi salah. Meskipun demikian, disadari bahwa program hasil modifikasi masih memiliki kekurangan, yaitu belum mampu dilampaui akurasi dari program *baseline*. Oleh karena itu, sebagai rencana untuk percobaan selanjutnya, akan dilakukan modifikasi pada *hyperparameter* LSTM dengan tujuan meningkatkan tingkat akurasi dan keakuratan hasil prediksi *gender* berdasarkan nama dalam bahasa Indonesia. Modifikasi yang direncanakan yaitu penyesuaian beberapa faktor, seperti ukuran *batch*, jumlah lapisan LSTM, jumlah unit LSTM, *learning rate*, fungsi *aktivasi*, dan *dropout rate* untuk meningkatkan pemahaman dan representasi fitur nama-nama dalam bahasa Indonesia yang beragam. Dengan demikian, diharapkan bahwa program yang telah dimodifikasi akan mengalami peningkatan yang signifikan dalam kemampuannya untuk memprediksi nama-nama, khususnya dalam konteks nama-nama bahasa Indonesia yang mencakup kombinasi dari nama laki-laki dan perempuan.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Ibu Alvi Syahrini Utami, M. Kom selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya dan Bapak Abdiansah selaku dosen pengampu yang telah memberikan bimbingan dan arahan serta memberikan saran pada proses penelitian ini. Dalam penulisan jurnal ilmiah ini tidak terlepas dukungan dari berbagai pihak. Terima kasih kepada semua pihak yang telah memberikan bimbingan, petunjuk, bantuan, serta dorongan dalam bentuk moral maupun material sehingga penulisan jurnal ini dapat berjalan dengan baik.

DAFTAR PUSTAKA

- [1] Y. Hu, C. Hu, T. Tran, T. Kasturi, E. Joseph and M. Gillingham, "What's in a name? – gender classification of names with character based machine learning models," *Data Mining and Knowledge Discovery*, vol. 35, no. 4, pp. 1537-1563, 2021.
- [2] I. Irfani, "GitHub - irfani/Jenis-Kelamin: Memprediksi jenis kelamin dari nama bahasa Indonesia menggunakan Machine Learning.," 2017. [Online]. Available: <https://github.com/irfani/Jenis-Kelamin/>.

- [3] N. Azzahra, "Implementasi deep learning untuk memprediksi jenis kelamin berdasarkan nama (studi kasus : nama-nama orang Indonesia)," [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/65563> .
- [4] A. Suganda Girsang, "Long Short Term Memory (LSTM)," 2019. [Online]. Available: <https://mti.binus.ac.id/2019/12/02/long-short-term-memory- lstm/>.
- [5] M. I. Sahnur, "Memprediksi jenis kelamin dari nama bahasa Indonesia menggunakan Machine Learning," 2017. [Online]. Available: <https://id.linkedin.com/pulse/memprediksi-jenis-kelamin-dari-nama-bahasa-indonesia-machine-sahnur/>.
- [6] R. D. W. Santosa, M. A. Bijaksana and A. Romadhony, "Implementasi Algoritma Long Short-term Memory (Lstm) Untuk Mendeteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia," *eProceedings of Engineering*, vol. 8, no. 1, 2021.
- [7] J. van de Weijer, G. Ren, J. van de Weijer, W. Wei and Y. Wang, "Gender identification in Chinese names," *Lingua*, vol. 234, p. 102759, 2020.
- [8] M. Fabien, "Maël Fabien," 2019. [Online]. Available: https://maelfabien.github.io/machinelearning/NLP_7/.
- [9] T. Hirano, "Gender Prediction from Name with LSTM," *Kaggle*, 2020.
- [10] R. C. B. Rego, V. M. L. Silva and V. M. Fernandes, "Predicting Gender by First Name Using Character-level Machine Learning," 2021. [Online]. Available: <https://arxiv.org/abs/2106.10156>.
- [11] M. H. Kabir, F. Ahmad, M. A. M. Hasan and J. Shin, "Gender Recognition of Bangla Names Using Deep Learning Approaches," *Applied Sciences*, vol. 13, no. 1, p. 522, 2022.
- [12] P. Kaushik, A. Gupta, P. P. Roy and D. P. Dogra, "EEG-Based Age and Gender Prediction Using Deep BLSTM-LSTM Network Model," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2634-2641, 2019.
- [13] S. Gajendran, M. D and V. Sugumaran, "Character level and word level embedding with bidirectional LSTM – Dynamic recurrent neural network for biomedical named entity recognition from literature," *Journal of Biomedical Informatics*, vol. 112, p. 103609, 2020.
- [14] H. Q. To, K. V. Nguyen, N. L.-T. Nguyen and A. G.-T. Nguyen, "Gender Prediction Based on Vietnamese Names with Machine Learning Techniques," *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, 2020.
- [15] O. Irawati and N. Nuraeni, "Optimasi Decision Tree Dengan Particle Swarm Optimization Untuk Klasifikasi Prediksi Gender Pengguna E-Commerce," *Journal of Information Engineering and Educational Technology*, vol. 6, no. 1, pp. 37-41, 2022.
- [16] A. Amao, "Machine Learning in Plain English: Building a Decision Tree Model to Classify Names by Gender: Part One," *Codementor*, 2017.
- [17] a. chugh, "Deep Learning Introduction to Long Short Term Memory," 2019. [Online]. Available: <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>.
- [18] A. A. Septiandri, "Predicting the Gender of Indonesian Names," 2017. [Online]. Available: <https://arxiv.org/abs/1707.07129>.
- [19] N. K. Manaswi, "RNN and LSTM," *Deep Learning with Applications Using Python*, pp. 115-126, 2018.
- [20] A. Saxena and T. Sukumar, "[PDF] Predicting bitcoin price using lstm And Compare its predictability with arima model," 2018. [Online]. Available: <https://www.semanticscholar.org/paper/Predicting-bitcoin-price-using-lstm-And-Compare-its-Saxena-Sukumar/4579ec623f4d0a2c4a0d0a817589fa3c2c3636c5>.
- [21] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68-75, 2017.
- [22] R. H. Assaad and S. Fayek, "Predicting the Price of Crude Oil and its Fluctuations Using Computational Econometrics: Deep Learning, LSTM, and Convolutional Neural Networks," *Econometric Research in Finance*, vol. 6, no. 2, pp. 119-137, 2021.
- [23] S. p. k. Renduchinthala, "Solving the Vanishing Gradient Problem with LSTMs," *LinkdIn*, 2023.
- [24] L. Zaman, S. Sumpeno and M. Hariadi, "Analisis Kinerja LSTM dan GRU sebagai Model Generatif untuk Tari Remo," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 2, p. 142, 2019.
- [25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [26] M. B. Herlambang, "Deep Learning: Recurrent Neural Networks - Page 4 of 4," 2019. [Online]. Available: <https://www.megabagus.id/deep-learning-recurrent-neural-networks/4/>.
- [27] G. E. Dahl, Dong Yu, Li Deng and A. Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30-42, 2012.
- [28] K. Eckhardt, "Predicting Gender from first names with an LSTM," 2018. [Online]. Available: <https://www.kaggle.com/code/karsteneckhardt/predicting-gender-from-first-names-with-an-lstm/>.
- [29] N. Kumar, "Classifying the Name Nationality of a Person using LSTM and Pytorch," 2019. [Online]. Available: <https://www.marktechpost.com/2019/12/18/classifying-the-name-nationality-of-a-person-using-lstm-and-pytorch>.
- [30] Ibm, "What is Naïve Bayes," 2023. [Online]. Available: <https://www.ibm.com/topics/naive-bayes/>.
- [31] Z. Zhang, "Naive Bayes Explained," 2019. [Online]. Available: <https://towardsdatascience.com/naive-bayes-explained-9d2b96f4a9c0/>.
- [32] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Machine Learning with Applications*, vol. 4, p. 100032, 2021.
- [33] ActiveWizards, "Sentiment Analysis with Naive Bayes and LSTM," 2023. [Online]. Available: <https://activewizards.com/blog/sentiment-analysis-with-naive-bayes-and-lstm/>.