

**METODE *BIDIRECTIONAL LONG SHORT-TERM MEMORY*
UNTUK MEMPREDIKSI PELUNASAN BIAYA HAJI
DI KEMENAG KOTA MALANG**

SKRIPSI

**Oleh :
NURJIHAN NABILAH RAMADHANI
NIM. 210605110061**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**METODE *BIDIRECTIONAL LONG SHORT-TERM MEMORY*
UNTUK MEMPREDIKSI PELUNASAN BIAYA HAJI
DI KEMENAG KOTA MALANG**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
NURJIHAN NABILAH RAMADHANI
NIM. 210605110061

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

**METODE *BIDIRECTIONAL LONG SHORT-TERM MEMORY*
UNTUK MEMPREDIKSI PELUNASAN BIAYA HAJI
DI KEMENAG KOTA MALANG**

SKRIPSI

**Oleh :
NURJIHAN NABILAH RAMADHANI
NIM. 210605110061**

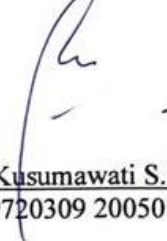
Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 5 Mei 2025

Pembimbing I,



Supriyono, M.Kom
NIP. 19841010 201903 1 012

Pembimbing II,



Dr. Ririen Kusumawati S.Si, M.Kom
NIP. 19720309 200501 2 002

Mengetahui,

Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Bachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

METODE *BIDIRECTIONAL LONG SHORT-TERM MEMORY* UNTUK MEMPREDIKSI PELUNASAN BIAYA HAJI DI KEMENAG KOTA MALANG

SKRIPSI

Oleh :
NURJIHAN NABILAH RAMADHANI
NIM. 210605110061

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 26 Mei 2025

Susunan Dewan Penguji

Ketua Penguji : Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

Anggota Penguji I : Nur Fitriyah Ayu Tunjung Sari, M.Cs
NIP. 19911226 202012 2 001

Anggota Penguji II : Supriyono, M.Kom
NIP. 19841010 201903 1 012

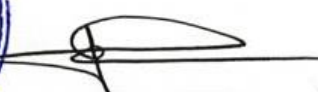
Anggota Penguji III : Dr. Ririen Kusumawati S.Si, M.Kom
NIP. 19720309 200501 2 002

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang




Bachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nurjihan Nabilah Ramadhani
NIM : 210605110061
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Metode *Bidirectional Long Short-Term Memory*
Untuk Memprediksi Pelunasan Biaya Haji Di
Kemenag Kota Malang

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 November 2025
Yang membuat pernyataan,



Nurjihan Nabilah Ramadhani
NIM. 210605110013

MOTTO

"Be bold in your dreams, be bold in your actions."

HALAMAN PERSEMBAHAN

Dengan rasa syukur kepada Allah SWT atas segala rahmat dan kemudahan-Nya, sehingga skripsi ini dapat terselesaikan dengan baik. Karya ini penulis persembahkan kepada Ayah Fauzi dan Ibu Yuni yang selalu mengalirkan kasih sayang, usaha terbaik, nasehat yang membangun, dukungan baik materi maupun non materi, dan do'a yang tiada henti. Kakak tersayang, M. Naufal Rafif Fikriyan dan Adek Tercinta Putra Maulana Rakha yang menjadi tempat berkeluh kesah dan dorongan untuk terus maju hingga skripsi ini dapat diselesaikan. Segenap keluarga besar, yang mengiringi perjalanan penulis dengan do'a dan dukungan. Aster Teknik Informatika angkatan 2021, yang telah berjuang bersama dari awal hingga akhir. Terakhir, kepada diri sendiri terima kasih karena sudah berusaha, bertahan, dan berjuang sejauh ini sampai semua terselesaikan.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur senantiasa penulis panjatkan kehadiran Allah SWT, atas berkat rahmat dan limpahan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Metode *Bidirectional Long Short-Term Memory* untuk Memprediksi Pelunasan Biaya Haji Di Kemenag Kota Malang” dengan baik dan lancar sebagai salah satu persyaratan untuk memperoleh gelar sarjana pada program studi Teknik Informatika jenjang Strata-1 Universitas Islam Negeri Maulana Malik Ibrahim Malang. Shalawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW, yang telah membawa kita dari zaman jahiliyah menuju zaman yang lurus yaitu Addinul Islam.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan motivasi kepada penulis sehingga dapat menyelesaikan skripsi ini. Ucapan terima kasih ini penulis disampaikan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

4. Supriyono, M.Kom selaku dosen pembimbing I yang sudah sabar dan mendukung penulis dengan memberikan bimbingan, motivasi, kritik dan saran untuk penulis dalam mengerjakan dan menyelesaikan skripsi ini.
5. Dr. Ririen Kusumawati S.Si, M.Kom selaku dosen pembimbing II yang sudah memberikan motivasi, mendukung dan membimbing penulis untuk mengerjakan dan menyelesaikan skripsi ini.
6. Dr. Zainal Abidin, M.Kom dan Nur Fitriyah Ayu Tanjung Sari, M.Cs selaku dosen penguji I dan II yang sudah memberikan masukan dan saran kepada penulis untuk menyelesaikan skripsi ini.
7. Kepada Ayahanda Fauzi Isdarto, cinta pertama penulis, terimakasih penulis ucapkan atas setiap kerja keras yang selama ini dilakukan sehingga putri mu ini bisa sampai kepada tahap ini, dan berkesempatan mengenyam pendidikan hingga bangku perkuliahan.
8. Kepada Ibu tersayang Sri Wahyuni, terimakasih karena beliau selalu mendukung, memotivasi dan mendoakan dalam mengejar masa depan putrinya sehingga penulis mampu menyelesaikan skripsi dan studi sampai sarjana
9. M. Naufal Rafif Fikriyan, kakak terkeren, yang selalu menjadi teladan, sumber inspirasi, dan motivasi terbesar penulis dalam menggapai cita-cita. Terima kasih atas dukungan, semangat, perhatian, dan semua nasihat berharga yang selalu menguatkan penulis dalam menjalani setiap proses hingga skripsi ini terselesaikan.

10. Putra Maulana Rakha Febrianto, adik tercinta saya yang paling tampan di keluarga besar, yang selalu menjadi tempat berbagi suka dan duka, mendengarkan setiap keluh kesah, serta memberikan dukungan, doa, dan semangat yang tulus. Terima kasih atas kehadiran dan kritikmu yang menjadi kekuatan tersendiri bagi penulis dalam menyelesaikan studi ini.
11. Kepada keluarga besar yang sudah memberikan do'a dan dukungan kepada penulis untuk tetap semangat menyelesaikan perkuliahan.
12. Nia Faricha S, Si selaku admin Program Studi Teknik Informatika yang sudah membantu penulis selama masa studi dan menyelesaikan skripsi ini.
13. Segenap dosen dan jajaran staff Program Studi Teknik Informatika yang telah memberikan ilmu dan pengetahuan selama penulis masa studi.
14. Sahabat terdekat penulis, Mutiara Aprilia, Salma Chesha, Sarah Aurelia, Aisya Gusti, Fitria Susanti, Frisca Yuni, Roudhlotul Hanna, Nabilah Mahdiyah, Anisa Puja, dan Amelia Amriadi yang sudah membantu dan mendukung penulis selama ini. Terima kasih atas kebersamaan, tawa, cerita, dan dukungan yang tulus dalam suka dan duka penulis selama masa perkuliahan hingga skripsi ini terselesaikan.
15. Terimakasih kepada Taylor Swift yang telah menemani penulis mengerjakan skripsi ini dengan lagu – lagunya dan *quotes* penyemangatnya *“Life can be heavy, especially if you try to carry it all at once. Part of growing up and moving into new chapters of your life is about catch and release “*

Penulis menyadari bahwa dalam penulisan skripsi masih terdapat kekurangan dan kesalahan. Penulis berharap semoga skripsi ini dapat memberikan manfaat kepada pembaca khususnya penulis sendiri.

Wassalamu'alaikum warahmatullahi wabarakatuh.

Malang, 20 Mei 2025

Penulis

DAFTAR ISI

HALAMAN PENGANTAR.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO	vi
HALAMAN PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xv
ABSTRAK	xvi
ABSTRAK	xvii
المخلص.....	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	4
BAB II STUDI PUSTAKA.....	5
2.1 Penelitian Terkait	5
2.2 Pelunasan Biaya Haji sebagai Objek Prediksi.....	7
2.3 <i>Bidirectional Long-Short term Memory (Bi-LSTM)</i>	10
BAB III METODE PENELITIAN	13
3.1 Desain Sistem	13
3.1.1 Arsitektur BI-LSTM	14
3.2 Pengumpulan Data	16
3.3 Data <i>Preprocessing</i>	20
3.4 Membangun Model BI-LSTM.....	22
3.4.1 Prediksi dengan BI-LSTM.....	24
3.5 Evaluasi Model.....	33
3.5.1 <i>Confusion Matrix</i>	33
3.5.2 <i>K-fold Cross Validation</i>	35
BAB IV HASIL DAN PEMBAHASAN.....	36
4.1 Data Input.....	36
4.1.1 Pembersihan Data.....	38
4.1.2 Pemilihan Fitur.....	41
4.1.3 Normalisasi Data	44
4.2 Pemodelan dan Pelatihan BI-LSTM.....	47
4.3 Pengujian BI-LSTM	47
4.4 Skenario Pengujian.....	49
4.5 Hasil.....	50
4.5.1 Hasil Pengujian Rasio 80:20.....	50
4.5.2 Hasil Pengujian Rasio 70:30.....	53
4.5.3 Hasil Evaluasi Model.....	56

4.6 Pembahasan.....	58
BAB V KESIMPULAN DAN SARAN	62
5.1 Kesimpulan.....	62
5.2 Saran.....	62
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1 Struktur gerbang LSTM	11
Gambar 3.1 Flowchart Proses Penelitian dengan Bi-LSTM.....	13
Gambar 3.2 Arsitektur BI-LSTM.....	14
Gambar 3.3 Skema partisi data dengan $K = 10$	35

DAFTAR TABEL

Tabel 3.1 Sampel Data	19
Tabel 3.2 Contoh data input yang telah dinormalisasi dan diencoding	24
Tabel 3.3 <i>Confusion Matrix</i>	33
Tabel 4.1 Data pelunasan biaya haji 2023-2024.....	37
Tabel 4.2 Pengecekan nilai yang hilang.....	38
Tabel 4.3 Hasil pengecekan tipe data.....	39
Tabel 4.4 Hasil proses <i>encoding</i>	41
Tabel 4.5 Fitur - fitur yang dipilih untuk prediksi	43
Tabel 4.6 Nilai minimum dan maksimum setiap fitur sebelum Normalisasi.....	44
Tabel 4.7 Hasil Normalisasi data	46
Tabel 4.8 Pembagian data	47
Tabel 4.9 Hasil <i>Train. Loss</i> dan <i>Val. Accuracy</i> 80:20 dengan 10-Fold.....	51
Tabel 4.10 Hasil 10-fold <i>Cross Validation</i> Rasio 80:20	53
Tabel 4.11 Hasil <i>Train. Loss</i> dan <i>Val. Accuracy</i> 70:30 dengan 10-Fold.....	53
Tabel 4.12 Hasil 10-fold <i>Cross Validation</i> Rasio 70:30	56
Tabel 4.13 Hasil Perbandingan Akurasi Kedua Skenario.....	57
Tabel 4.14 Perbandingan Prediksi Pelunasan Haji	57
Tabel 4.15 Hasil <i>confusion matrix</i> model terbaik.....	58

ABSTRAK

Ramadhani, Nurjihan Nabilah. 2025. **Metode *Bidirectional Long Short-Term Memory* untuk Memprediksi Pelunasan Biaya Haji di KEMENAG Kota Malang.** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Supriyono, M.Kom. (II) Dr. Ririen Kusumawati, S.Si, M.Kom

Kata Kunci: Bi-LSTM, Klasifikasi *Biner*, Prediksi, Pelunasan Biaya Haji

Prediksi pelunasan biaya haji menjadi kebutuhan penting dalam pengelolaan kuota dan pelayanan calon jemaah haji. Penelitian ini bertujuan untuk membangun model prediksi pelunasan haji menggunakan metode *Bidirectional Long Short-Term Memory* (Bi-LSTM) dengan output klasifikasi *biner*: melunasi atau tidak melunasi. Data diperoleh dari Kantor Kementerian Agama Kota Malang, mencakup fitur-fitur seperti usia, jenis kelamin, pendidikan, pekerjaan, bulan dan tahun daftar, serta bulan dan tahun pelunasan. Setelah melalui tahapan *preprocessing* berupa *encoding*, normalisasi, dan pembagian data (80:20 dan 70:30), kemudian model dilatih dengan pendekatan *10-Fold Cross Validation* untuk menghindari *overfitting* dan menguji konsistensi performa model. Hasil terbaik diperoleh pada rasio 70:30 dengan akurasi sebesar 83,04% pada *fold* ke-7 hal ini menunjukkan bahwa model mampu mengenali pola temporal antar fitur dan menghasilkan klasifikasi yang akurat dan seimbang. Penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan sistem prediksi berbasis kecerdasan buatan dalam layanan haji di Indonesia.

ABSTRAK

Ramadhani, Nurjihan Nabilah. 2025. ***Bidirectional Long Short-Term Memory Method for Predicting Hajj Payment Completion at the Ministry of Religious Affairs in Malang City***. Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Supriyono, M.Kom., (II) Dr. Ririen Kusumawati, M.Kom.

Predicting hajj payment settlement is an important need in managing quotas and services for prospective hajj pilgrims. This study aims to develop a prediction model for hajj payment settlement using the Bidirectional Long Short-Term Memory (Bi-LSTM) method with a binary classification output: paid or not paid. Data was obtained from the Ministry of Religious Affairs Office in Malang City, which includes features such as age, gender, education, occupation, registration month and year, as well as payment settlement month and year. After undergoing preprocessing stages, including encoding, normalization, and data splitting (80:20 and 70:30), the model was trained using a 10-Fold Cross Validation approach to avoid overfitting and test the model's performance consistency. The best results were achieved with a 70:30 ratio, with an accuracy of 83.04% on fold 7, indicating that the model was able to recognize temporal patterns across features and generate accurate and balanced classifications. This research is expected to contribute scientifically to the development of artificial intelligence-based prediction systems in hajj services in Indonesia.

Key words: Bi-LSTM, Binary Classification, Prediction, Hajj Payment Settlement

الملخص

رمضاني، نورجيهان نبيلة. 2025. طريقة الذاكرة طويلة المدى ثنائية الاتجاه للتنبؤ بتسوية تكاليف الحج في مدينة كيميناچ مالانچ. أطروحة. برنامج دراسات هندسة المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا، مالك إبراهيم الإسلامية الحكومية مالانچ. المشرفون: (أ) سوبريونو، م. كوم. (ب) د. ريرين كوسوماوتي. س. سي، م. كوم.

سداد رسوم الحج، وزارة الشؤون الدينية، التنبؤ، التصنيف الثنائي، التعلم، Bi-LSTM: الكلمات المفتاحية K-Fold Cross Validation العميق، التحقق المتقاطع

يعد التنبؤ بسداد رسوم الحج مطلباً مهماً في إدارة الحصص وخدمة الحجاج المحتملين. يهدف هذا البحث إلى بناء نموذج للتنبؤ بسداد رسوم الحج باستخدام طريقة الذاكرة ثنائية الاتجاه قصيرة المدى مع مخرجات تصنيف ثنائي: السداد أو عدم السداد. وقد تم الحصول (Bi-LSTM) طويلة الأجل ثنائية الاتجاه على البيانات من مكتب مدينة مالانچ التابع لوزارة الشؤون الدينية، بما في ذلك ميزات مثل العمر والجنس والتعليم والمهنة وشهر وسنة التسجيل وشهر وسنة السداد. بعد المرور بمراحل ما قبل المعالجة من ترميز والتحقق المتقاطع *K-Fold Cross Validation* وتطبيق ومشاركة البيانات، تم تدريب النموذج باستخدام نهج التحقق المتقاطع لتجنب الإفراط في التهيئة واختبار اتساق أداء النموذج. تم تدريب النموذج باستخدام سيناريوهين لنسبة البيانات واختباره باستخدام نهج التحقق المتقاطع 10 أضعاف لتجنب الإفراط في التركيب. تم (و70:30 80:20) الحصول على أفضل النتائج عند نسبة 70:30 مع دقة 83.04% في الطية السابعة. أثبت النموذج قدرته على التعرف على الأنماط الزمنية بين السمات وإنتاج تصنيفات دقيقة ومتوازنة. من المتوقع أن يقدم هذا البحث مساهمة علمية في تطوير أنظمة التنبؤ القائمة على الذكاء الاصطناعي في خدمات الحج في إندونيسيا.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pelaksanaan ibadah haji merupakan salah satu rukun Islam yang memiliki peran penting bagi umat Muslim. Ibadah haji merupakan salah satu rukun Islam yang wajib dilaksanakan oleh umat Muslim yang mampu secara finansial dan fisik. Pelaksanaannya melibatkan berbagai tahapan, mulai dari pendaftaran, pelunasan biaya, hingga keberangkatan jamaah. Di Indonesia, pelaksanaan ibadah haji dikelola oleh Kementerian Agama (Kemenag), yang bertanggung jawab memastikan bahwa setiap proses berjalan dengan lancar. Salah satu tahapan krusial dalam penyelenggaraan ibadah haji adalah pelunasan Biaya Penyelenggaraan Ibadah Haji (BPIH), yang harus dilakukan oleh calon jamaah sebelum keberangkatan. Dengan jumlah jamaah yang terus meningkat setiap tahunnya, pengelolaan data pelunasan menghadirkan tantangan yang signifikan, baik dari sisi volume data maupun variasi pembayaran (Qolbi & Rahman, 2023).

Setiap calon jamaah yang mendaftar akan mendapatkan nomor porsi dan masuk ke dalam daftar tunggu berdasarkan prinsip *first come, first served*. Setelah mendapat kesempatan berangkat, calon jamaah diwajibkan melakukan pelunasan Biaya Penyelenggaraan Ibadah Haji (Bipih) sebagai salah satu syarat keberangkatan. Namun dalam pelaksanaannya, tidak semua calon jamaah yang dipanggil melakukan pelunasan secara tepat waktu. Beberapa dari mereka justru memilih tidak melunasi. Ketidakteraturan dalam proses pelunasan ini dapat

disebabkan oleh berbagai faktor, seperti kondisi ekonomi yang belum stabil, pertimbangan kesehatan, perubahan rencana pribadi, maupun kendala administratif lainnya. Akibatnya, proses penyerapan kuota haji menjadi kurang optimal dan dapat berdampak pada efisiensi perencanaan serta pengelolaan pelayanan oleh pihak penyelenggara. Dalam jangka panjang, ketidakpastian ini berpotensi memengaruhi sistem distribusi kuota secara nasional dan efektivitas keberangkatan jemaah secara menyeluruh.

Untuk menghadapi tantangan tersebut, Kementerian Agama telah menggunakan Sistem Komputerisasi Haji Terpadu (SISKOHAT) sebagai alat untuk mengelola data pelunasan secara lebih terstruktur. SISKOHAT dirancang untuk mencatat data secara real-time, mulai dari proses pendaftaran hingga pembayaran pelunasan oleh jamaah. Namun, sistem ini masih berfungsi sebagai alat administrasi dan belum memiliki kemampuan analisis prediktif untuk mempelajari pelunasan di masa mendatang. Padahal, pemahaman mengenai pelunasan dapat memberikan manfaat besar bagi perencanaan pengelolaan administrasi haji (Kusuma & Hadinata, 2022).

Dalam penelitian ini, metode Bi-LSTM diterapkan untuk menganalisis data pelunasan biaya haji. Bi-LSTM merupakan salah satu teknik dalam *machine learning* yang dirancang untuk menangani data berbasis waktu (*time series*). Keunggulan metode ini adalah kemampuannya untuk memproses data dalam dua arah, yaitu dari masa lalu ke masa depan dan sebaliknya, sehingga dapat menangkap temporal yang kompleks (Ajayaghosh, 2024). Dengan memanfaatkan

data historis pelunasan, model Bi-LSTM dapat mengenali pembayaran yang telah terjadi dan memberikan prediksi pelunasan di masa depan.

Penelitian ini difokuskan pada analisis pelunasan biaya haji di Kemenag Kota Malang menggunakan data historis pelunasan yang telah tercatat. Melalui penerapan Bi-LSTM, penelitian ini menghasilkan model prediktif yang mampu memberikan gambaran tentang kemungkinan jumlah calon jamaah yang akan melunasi biaya haji dalam jangka waktu tertentu.

Kebaruan penelitian ini terletak pada penerapan Bi-LSTM dalam konteks pelunasan biaya haji di Indonesia, khususnya di Kemenag Kota Malang. Hingga saat ini, metode Bi-LSTM lebih sering digunakan untuk analisis data dalam bidang lain, seperti keuangan atau prediksi cuaca. Penerapan metode ini untuk menganalisis data pelunasan biaya haji menjadi langkah inovatif dalam memanfaatkan teknologi machine learning untuk mendukung pengelolaan administrasi haji.

Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam bidang pendidikan, khususnya dalam pengembangan teknologi untuk mempermudah proses administrasi dan manajemen data yang terkait dengan biaya haji.

1.2 Rumusan Masalah

Permasalahan yang menjadi fokus dalam penelitian ini adalah mengenai bagaimana model (Bi-LSTM) dapat digunakan untuk memprediksi pelunasan biaya haji di masa mendatang berdasarkan data historis yang tersedia.

1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah mengimplementasikan metode Bi-LSTM untuk memprediksi pelunasan biaya haji Kemenag Kota Malang.

1.4 Batasan Masalah

Adapun batasan yang ada pada penelitian ini adalah sebagai berikut:

- a. Penelitian menggunakan data pelunasan biaya haji yang dengan rentang tahun 2023 – 2024 Kemenag Kota Malang.
- b. Penelitian ini hanya mengevaluasi akurasi prediksi pelunasan biaya haji dengan menggunakan matriks *precision*, *recall*, dan *F1-score*. tanpa membandingkan dengan model lain.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam menjawab pertanyaan sejauh mana penerapan model Bi-LSTM dapat memberikan kontribusi terhadap pengembangan pengetahuan dan pemahaman di bidang prediksi pelunasan biaya haji.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian terkait merupakan bagian penting dalam menyusun studi pustaka dan memberikan konteks bagi penelitian yang sedang dilakukan. Terdapat beberapa penelitian telah menerapkan metode Bi-LSTM untuk memproses data berurutan dan menghasilkan prediksi dua kondisi. Kajian ini menjadi dasar dalam mendukung pemilihan metode pada penelitian ini.

Pada penelitian oleh Binbusayyis & Sha (2025), dikembangkan model *deep learning* berbasis CNN-BiLSTM yang ditujukan untuk memprediksi konsumsi energi rumah tangga melalui data deret waktu. Dengan memanfaatkan kemampuan Bi-LSTM dalam mengenali pola urutan data secara dua arah serta ditunjang oleh mekanisme perhatian (*attention*), model ini menghasilkan prediksi dalam dua kategori, yaitu “tinggi” dan “rendah”. Evaluasi model menunjukkan tingkat akurasi mencapai 90%, yang menegaskan efektivitas arsitektur ini dalam menangani data *time-series* untuk tujuan prediktif.

Pada penelitian oleh Thekkekara et al. (2024), metode CNN-BiLSTM digunakan untuk memprediksi kecenderungan depresi pada pengguna media sosial berbasis teks. Sistem ini menghasilkan dua kemungkinan prediksi, yakni “depresi” dan “tidak depresi”. Dengan dukungan *attention mechanism* yang memperkuat penangkapan konteks emosional, model ini mampu mencapai akurasi 96,7%,

menunjukkan performa tinggi dalam memahami data linguistik untuk tujuan prediksi kondisi psikologis.

Pada penelitian oleh K et al. (2024), dikembangkan model prediksi terhadap keberadaan rumor dalam teks media sosial. Pendekatan ini menggabungkan metode embedding BERT-OPCNN dan FastText-FIAC, yang kemudian diproses menggunakan Bi-LSTM berbasis pembobotan kelas (*cost-sensitive*). Hasilnya menunjukkan akurasi sebesar 98,24%, membuktikan keandalan model ini dalam memprediksi keberadaan informasi yang bersifat benar atau tidak.

Pada penelitian oleh N. et al. (2023), dilakukan pengembangan sistem penilaian otomatis menggunakan CNN dan Bi-LSTM untuk memprediksi kelayakan tulisan tangan peserta didik dalam ujian. Model ini memberikan hasil berupa dua prediksi: “layak” dan “tidak layak”. Dengan akurasi mencapai 95,26%, penelitian ini menunjukkan bahwa pendekatan Bi-LSTM mampu mengolah data tekstual berurutan dalam konteks evaluasi pendidikan secara efektif.

Pada penelitian oleh Hu et al. (2020), diajukan metode Bi-LSTM untuk memprediksi kondisi kejang pada pasien berdasarkan sinyal EEG kulit kepala. Dengan dukungan teknik *Local Mean Decomposition* (LMD) sebagai praproses data, model ini mampu mengidentifikasi dua kondisi: “kejang” dan “non-kejang”. Hasil pengujian menunjukkan sensitivitas 93,61%, spesifisitas 91,85%, dan nilai G-Mean sebesar 92,66%, menandakan kemampuan model dalam membuat prediksi medis berbasis sinyal biologis secara akurat.

Tabel 2.1 menyajikan beberapa penelitian terdahulu yang menggunakan metode Bi-LSTM untuk memprediksi kondisi dua arah atau dua hasil berbeda. Pada

penelitian-penelitian tersebut, data yang digunakan bersifat berurutan dan dianalisis untuk menghasilkan keputusan prediktif. Berikut penjelasan singkat penelitian-penelitian sebelumnya:

Tabel 2.1 Penelitian Terkait

Referensi	Topik	Metode	Hasil/Akurasi
(Binbusayyis & Sha, 2025)	Prediksi konsumsi energi menggunakan CNN-BiLSTM.	BiLSTM	90%
(Philip Thekkekara et al., 2024)	Prediksi untuk mendeteksi depresi berdasarkan teks pengguna	CNN-BiLSTM	96,7%
(K et al., 2024)	Prediksi klasifikasi untuk mendeteksi rumor berbasis teks di media sosial	BiLSTM	98.24%
(N. et al., 2023)	Prediksi berbasis kecocokan/sekuens untuk evaluasi otomatis tulisan tangan (jawaban ujian)	CNN dan Bi-LSTM	95.26%
(Hu et al., 2020)	Prediksi untuk mendeteksi kejang epilepsi dari sinyal EEG scalp.	Bi-LSTM	92.66%,

2.2 Pelunasan Biaya Haji sebagai Objek Prediksi

Pelunasan biaya haji merupakan tahap administratif yang sangat penting dalam proses pemberangkatan jemaah haji di Indonesia. Setelah melakukan pendaftaran awal dan memperoleh nomor porsi, setiap calon jemaah diwajibkan melunasi Biaya Penyelenggaraan Ibadah Haji (Bipih) sebagai syarat mutlak agar dapat diberangkatkan ke Tanah Suci. Proses pelunasan ini dicatat secara sistematis dalam database nasional milik Kementerian Agama, termasuk pada tingkat daerah seperti Kantor Kemenag Kota Malang.

Dalam konteks ini, pelunasan biaya haji dapat dilihat sebagai bentuk keputusan administratif yang dipengaruhi oleh sejumlah variabel sosial, ekonomi, dan kebijakan. Salah satu faktor utama adalah usia calon jemaah. Usia lanjut, terutama di atas 80 tahun, sering kali mendapat prioritas keberangkatan berdasarkan

kebijakan pemerintah. Dengan demikian, jemaah dalam kategori usia ini cenderung melakukan pelunasan lebih cepat ketika kesempatan diberikan. Selain itu, kondisi fisik dan administratif pada kelompok usia tertentu juga dapat memengaruhi kelancaran proses pelunasan (Islami Rahayu et al., 2024).

Faktor lain yang signifikan adalah pekerjaan dan penghasilan. Tingkat ekonomi seseorang secara langsung memengaruhi kemampuannya dalam melunasi biaya perjalanan haji. Calon jemaah dengan penghasilan tetap dan stabil cenderung memiliki kesiapan finansial yang lebih tinggi dibandingkan dengan mereka yang berpenghasilan tidak menentu. Dalam hal ini, pekerjaan berfungsi sebagai indikator ekonomi yang dapat digunakan untuk memperkirakan potensi pelunasan (Nyoman, 2020).

Lama masa tunggu juga menjadi faktor yang tidak bisa diabaikan. Dengan waktu tunggu yang bisa mencapai belasan hingga puluhan tahun, kondisi calon jemaah bisa mengalami perubahan dari segi ekonomi, kesehatan, maupun semangat untuk berangkat (Wardatur Rizqiyah & Chonyta, 2024). Perubahan-perubahan ini dapat menyebabkan ketidakteraturan dalam pelunasan meskipun jemaah telah masuk ke daftar keberangkatan. Selain itu, tahun pendaftaran juga memiliki pengaruh tersendiri. Semakin awal tahun pendaftaran, maka semakin besar kemungkinan jemaah masuk dalam kuota lebih awal dan melakukan pelunasan lebih cepat, karena waktu estimasi keberangkatan lebih pasti.

Jenis pelunasan yang dilakukan juga menjadi indikator penting. Jemaah reguler memiliki peluang pelunasan yang lebih stabil karena kuota mereka telah ditetapkan sejak awal. Sebaliknya, jemaah cadangan berada dalam posisi yang tidak

pasti karena hanya akan diberangkatkan apabila terdapat sisa kuota dari pelunasan tahap pertama (Sofianiyatin & Pradana, 2024). Hal ini menyebabkan ketidakteraturan waktu pelunasan dari jemaah cadangan, yang bisa memengaruhi pola data.

Selain itu, faktor kebijakan pemerintah juga turut memengaruhi. Setiap tahun, pemerintah melalui Kementerian Agama dan BPKH menetapkan besaran Bipih berdasarkan berbagai pertimbangan, termasuk fluktuasi kurs dan biaya layanan di Arab Saudi (Anas et al., 2023). Penyesuaian biaya ini dapat mendorong sebagian jemaah menunda pelunasan jika tidak mampu menyesuaikan kondisi finansialnya secara cepat.

Dengan mempertimbangkan keseluruhan faktor, pelunasan biaya haji tidak hanya merupakan proses administratif biasa, melainkan juga merupakan fenomena yang kompleks dan dinamis. Maka dari itu, menjadikan pelunasan biaya haji sebagai objek prediksi merupakan langkah strategis yang sangat relevan. Melalui pendekatan analitik berbasis data historis, hasil prediksi dari model seperti Bi-LSTM dapat memberikan kontribusi dalam pengembangan sistem pengambilan keputusan berbasis data (*data-driven decision making*), yang semakin dibutuhkan dalam tata kelola pelayanan publik modern.

Dalam penelitian ini, pendekatan Bi-LSTM digunakan untuk memproses data pelunasan biaya haji dan menghasilkan prediksi status pelunasan calon Jemaah, apakah mereka diperkirakan akan melunasi atau tidak berdasarkan data historis yang tersedia. Hasil dari prediksi ini diharapkan dapat menjadi landasan awal dalam pengembangan sistem informasi yang lebih adaptif, tetapi juga mendukung upaya

perencanaan kebijakan yang lebih responsif dan efisien di lingkungan Kementerian Agama Kota Malang.

2.3 *Bidirectional Long-Short term Memory (Bi-LSTM)*

Bidirectional Long Short-Term Memory (Bi-LSTM) merupakan pengembangan dari arsitektur *Long Short-Term Memory (LSTM)*, yang pada dasarnya adalah salah satu varian dari *Recurrent Neural Network (RNN)*. Bi-LSTM dirancang untuk mengatasi kelemahan RNN tradisional dalam mengolah data berurutan, khususnya terkait dengan masalah *vanishing gradient* dan ketidakmampuan dalam menangkap dependensi jangka panjang (Zen et al., 2024). Bi-LSTM merupakan metode yang sering digunakan untuk pemrosesan urutan data, terutama dalam tugas-tugas yang membutuhkan pemahaman konteks dari dua arah, baik dari masa lalu maupun masa depan (Tran et al., 2021).

LSTM memiliki struktur khusus berupa *memory cell* yang dapat menyimpan informasi dalam periode waktu yang panjang melalui tiga gerbang utama, yaitu *input gate*, *forget gate*, dan *output gate*. Arsitektur ini memungkinkan model untuk secara selektif menyimpan, melupakan, atau mengeluarkan informasi pada setiap langkah waktu, menjadikannya sangat efektif dalam memproses data sekuensial.

Fungsi Aktivasi adalah fungsi yang digunakan pada RNN yang berfungsi untuk menentukan bagaimana informasi ditransformasikan dalam unit-unit neuron di dalam jaringan pada setiap langkah waktu (Justin et al., 2024). Pada gambar 2.1 fungsi aktivasi yang digunakan dalam LSTM adalah fungsi *Tanh* dan *Sigmoid*. Fungsi aktivasi *Tanh* digunakan untuk memperbarui hidden state jaringan dengan cara mengatur informasi yang relevan yang dibawa ke waktu berikutnya. Fungsi

aktivasi *Tanh* mengubah nilai input menjadi output dalam rentang antara -1 dan 1 (Lasijan et al., 2023). Persamaan dari fungsi *Tanh* dapat dilihat dari persamaan (2.1).

$$\tanh(x) = \frac{e^{\{2x\}} - 1}{e^{\{2x\}} + 1} \quad (2.1)$$

Keterangan :

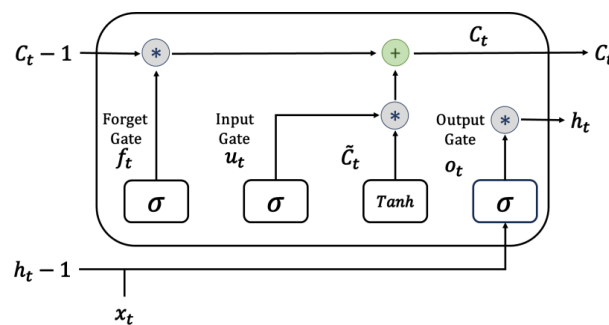
- a) $\tanh(x)$ = Hasil dari fungsi hyperbolic tangent untuk nilai x
- b) x = Input
- c) e = Euler

Fungsi aktivasi *Sigmoid* digunakan pada *Forget Gate*, *Input Gate*, dan *Output Gate*. Fungsi sigmoid memetakan input ke dalam rentang 0 dan 1, yang memungkinkan model untuk memutuskan apakah suatu informasi akan dilupakan atau disimpan (Lasijan et al., 2023). Persamaan dari fungsi *Tanh* dapat dilihat dari persamaan (2.2).

$$\sigma(x) = \frac{1}{1 + e^{\{-x\}}} \quad (2.2)$$

Keterangan :

- a) $\sigma(x)$: Sigmoid
- b) x : Input
- c) e : Euler



Gambar 2.1 Struktur gerbang LSTM

Bi-LSTM memperluas kemampuan LSTM dengan menggabungkan dua lapisan LSTM yang berjalan secara paralel namun dengan arah yang berbeda. Lapisan pertama memproses data dari awal hingga akhir (*forward*), sedangkan lapisan kedua memproses data dari akhir hingga awal (*backward*). Penggabungan ini dilakukan melalui operasi *concatenation*, di mana hasil dari kedua LSTM tersebut digabungkan menjadi satu vektor panjang. Penggabungan output dalam Bi-LSTM dapat dinyatakan dalam persamaan 2.6 berikut :

$$y_t = (\overrightarrow{h_t}; \overleftarrow{h_t}) \quad (2.6)$$

Keterangan :

- a) y_t : vektor gabungan
- b) $;$: fungsi *concatenation*
- c) $\overrightarrow{h_t}$: *hidden states* pada *forward pass*
- d) $\overleftarrow{h_t}$: *hidden states* pada *backward pass*

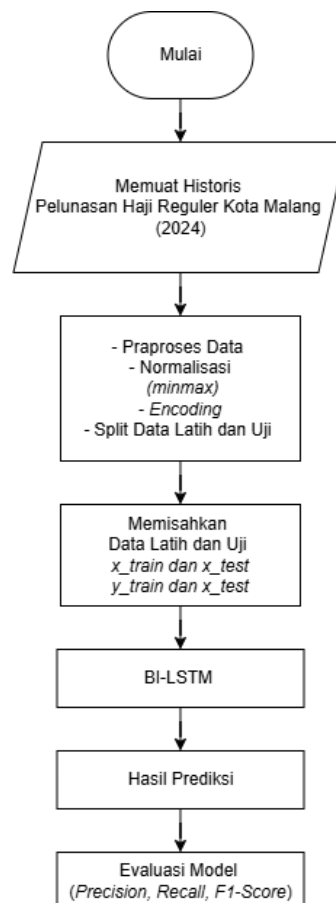
Penggabungan kedua arah ini memungkinkan model untuk mempelajari pola yang lebih kompleks dan memahami konteks data dengan lebih baik dibandingkan dengan LSTM.

BAB III

METODE PENELITIAN

3.1 Desain Sistem

Desain sistem pada penelitian ini menggambarkan alur kerja utama yang dilakukan dalam proses implementasi model. Gambar 3.1 memperlihatkan tahapan-tahapan penelitian yang berfokus pada penerapan metode Bi-LSTM untuk memprediksi pelunasan biaya haji. Seluruh proses ditampilkan dalam bentuk *flowchart*.

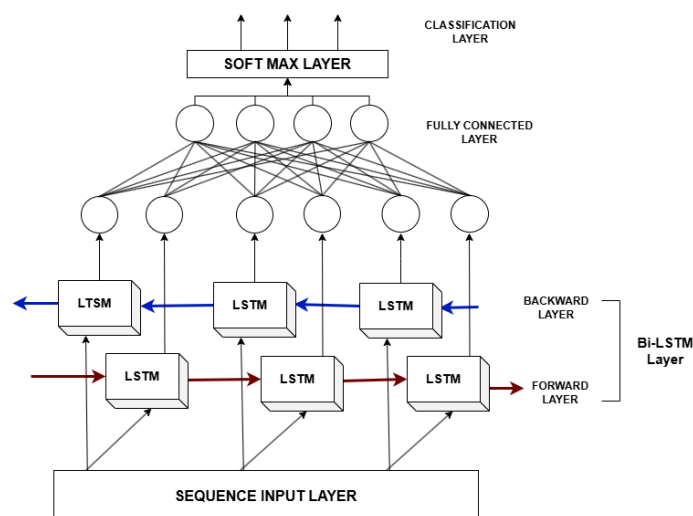


Gambar 3.1 Flowchart Proses Penelitian dengan Bi-LSTM

Flowchart Gambar 3.1 menjelaskan alur penelitian menggunakan pendekatan Bi-LSTM. Tahapan dimulai dari proses Input Data, yaitu pengumpulan data historis pelunasan calon jemaah haji yang diambil dari Kementerian Agama Kota Malang. Data yang dikumpulkan terdiri dari variabel-variabel yang dianggap relevan terhadap proses prediksi pelunasan, antara lain: Nomor Porsi, Jenis Kelamin, Umur, Pendidikan, Pekerjaan, Bulan Daftar, Tahun Daftar, Bulan Lunas, Tahun Lunas, dan Jangka Waktu Pelunasan.

3.1.1 Arsitektur BI-LSTM

Arsitektur Bi-LSTM merupakan pengembangan dari LSTM yang dirancang untuk menangkap pola hubungan dalam data sekuensial secara lebih komprehensif. Tidak seperti LSTM biasa yang hanya memproses informasi dalam satu arah (*forward*), Bi-LSTM bekerja dengan dua jalur pemrosesan paralel, *forward layer* dan *backward layer*.



Gambar 3.2 Arsitektur BI-LSTM

Pada arsitektur yang ditampilkan dalam Gambar 3.2, proses dimulai dari *Sequence Input Layer* yang berfungsi menerima input data berurutan dalam bentuk vektor numerik. Data ini merupakan hasil dari tahap praproses, seperti normalisasi untuk menyetarakan skala antar fitur serta encoding pada data kategorik agar dapat diolah oleh model. Input yang telah dipersiapkan tersebut kemudian diproses secara simultan oleh dua jalur utama dalam struktur Bi-LSTM, yaitu *forward layer* dan *backward layer*. Jalur *forward* (ditandai dengan panah merah) memproses data dari urutan awal ke akhir, sedangkan jalur *backward* (ditandai dengan panah biru) memproses data dari akhir ke awal. Kedua jalur ini terdiri dari unit-unit LSTM yang masing-masing dilengkapi dengan tiga gerbang utama, yaitu *forget gate*, *input gate*, dan *output gate*. Gerbang-gerbang ini bertugas mengatur aliran informasi di dalam setiap unit, mulai dari menyimpan informasi penting, membuang informasi yang tidak relevan, hingga menghasilkan representasi internal yang baru. *Output* dari kedua jalur Bi-LSTM tersebut kemudian digabungkan dan diteruskan ke *Fully Connected Layer*, yaitu lapisan yang berfungsi mengintegrasikan hasil pembelajaran menjadi representasi padat yang siap dianalisis lebih lanjut. Selanjutnya, hasil dari lapisan ini dikirim ke *Softmax Layer* sebagai *classification layer* untuk menghitung probabilitas dari masing-masing kelas keluaran. Dalam konteks penelitian ini, *softmax* digunakan untuk memutuskan hasil akhir prediksi, yakni apakah seorang calon jemaah diperkirakan “Melunasi” atau “Tidak Melunasi” dalam pelunasan biaya haji.

Secara keseluruhan, struktur Bi-LSTM ini dinilai sangat tepat digunakan dalam pengolahan data administratif berurutan karena kemampuannya dalam

mengenali pola jangka panjang dan menangkap dinamika temporal dari dua arah waktu sekaligus. Dengan keunggulan tersebut, arsitektur ini diharapkan mampu memberikan hasil prediksi yang lebih akurat dalam memproyeksikan pelunasan biaya haji berdasarkan data historis yang tersedia.

3.2 Pengumpulan Data

Adapun data yang dimanfaatkan dalam penelitian ini berasal dari data pelunasan biaya haji tahun 2023-2024, dengan tahun pendaftaran jemaah yang tercatat sejak 2008 hingga 2019 dengan total 1.619 entri dan 12 fitur atau atribut data yang beragam. Dari keseluruhan atribut tersebut, sebanyak 9 parameter digunakan sebagai variabel input dalam proses pemodelan, yaitu jenis kelamin, umur, pendidikan, pekerjaan, bulan daftar, tahun daftar, bulan lunas, tahun lunas, dan jangka waktu pelunasan. Sementara itu, atribut status pelunasan digunakan sebagai target yang akan diprediksi oleh model. Secara keseluruhan, jumlah data yang digunakan dalam penelitian ini berjumlah 1.619 entri.

Parameter yang digunakan dalam penelitian ini dipilih berdasarkan keterkaitannya yang langsung dengan perilaku pelunasan biaya haji serta kontribusinya dalam membentuk prediksi pelunasan biaya haji. Pertama adalah jenis kelamin, karena perbedaan peran dan tanggung jawab sosial antara laki-laki dan perempuan dapat memengaruhi prioritas serta kemampuan dalam menyelesaikan pembayaran. Peran perempuan dalam konteks haji sering kali berbeda dalam hal prioritas dan akses pembiayaan. Selain itu, laki-laki yang umumnya menjadi kepala keluarga mungkin memiliki kecenderungan melunasi

lebih awal karena tanggung jawab keuangan yang lebih besar (El-Kafrawy et al., 2023).

Kedua, Umur menjadi parameter penting karena calon jemaah berusia lanjut berpotensi mendapatkan prioritas keberangkatan sesuai kebijakan Kementerian Agama yang tertuang dalam Peraturan Menteri Agama (PMA) No. 13 Tahun 2021. Artinya, jemaah lansia cenderung lebih cepat dalam menyelesaikan pelunasan saat dipanggil (UJCS, 2021). Oleh karena itu, usia bukan sekadar data demografis, melainkan juga indikator prioritas pelayanan.

Ketiga, pendidikan berperan dalam menentukan seberapa baik calon jemaah memahami prosedur administrasi pelunasan. Individu dengan tingkat pendidikan tinggi cenderung lebih siap secara administratif, sehingga dapat menyelesaikan proses pelunasan lebih awal (Mauludi, 2022). Hal ini berkaitan erat dengan pekerjaan, karena jenis pekerjaan merepresentasikan kestabilan ekonomi. Pekerjaan formal seperti PNS atau pegawai swasta sering kali diiringi dengan pendapatan tetap, yang tentunya memengaruhi kesiapan finansial untuk melunasi biaya haji (Syahbilqis, 2022).

Keempat, Bulan dan Tahun Daftar berkorelasi langsung dengan waktu panggilan pelunasan. Jemaah yang mendaftar lebih awal, khususnya sebelum tahun 2012, mayoritas telah mendapat jadwal pelunasan tahun 2023 dan 2024. Namun, perbedaan bulan pendaftaran dalam satu tahun dapat menyebabkan perbedaan bulan pelunasan, yang menunjukkan bahwa sistem pemanggilan sangat memperhatikan urutan waktu secara spesifik. Selain itu, Bulan dan Tahun Lunas menjadi indikator aktual dari kapan pelunasan dilakukan, dan ketika dibandingkan dengan data

pendaftaran, dapat dihitung Jangka Waktu Pelunasan, yaitu lamanya waktu antara pendaftaran dan pelunasan. Dalam banyak kasus, jangka waktu ini berkisar antara 11 hingga 16 tahun. Menariknya, terdapat jemaah yang memiliki latar belakang demografis serupa, namun memiliki status pelunasan yang berbeda, yang mengindikasikan adanya faktor-faktor psikologis, ekonomi musiman, atau kebijakan pemerintah yang ikut memengaruhi (Anas et al., 2023).

Oleh karena itu, semua parameter yang digunakan dalam penelitian ini saling berkaitan dan saling memengaruhi dalam proses pelunasan biaya haji. Melalui model prediksi berbasis Bi-LSTM, relasi antar parameter tersebut dapat dianalisis secara lebih mendalam dan menyeluruh.

Berdasarkan diagram alur desain sistem pada Gambar 3.1, tahap pertama dalam penelitian ini adalah pengumpulan data. Data yang digunakan merupakan data primer yang diperoleh langsung dari Kantor Kementerian Agama (Kemenag) Kota Malang.

Tabel 3.1 Sampel Data

Porsi	Jenis Kelamin	Umur	Pendidikan	Pekerjaan	Bulan Daftar	Tahun Daftar	Bulan Lunas	Tahun Lunas	Jangka Waktu Pelunasan	Status Pelunasan
1300502734	L	60	S1	Pegawai Negri Sipil	1	2011	2	2023	12	Melunasi
1300590684	L	56	S1	Swasta	12	2011	1	2023	12	Melunasi
1300590686	P	55	S1	Swasta	12	2011	1	2023	12	Melunasi
1300575727	P	62	SLTA	Ibu Rumah Tangga	12	2011	1	2023	12	Tidak Melunasi
1300590920	L	29	SD	Swasta	12	2011	2	2023	12	Melunasi

3.3 Data Preprocessing

Tahap *preprocessing* data merupakan bagian penting dalam proses pemodelan, karena kualitas dan kesiapan data akan sangat menentukan performa model prediksi yang dibangun. Dalam penelitian ini, *preprocessing* dilakukan secara bertahap mulai dari pemuatan data mentah hingga tahap pemisahan fitur dan label yang siap digunakan untuk pelatihan model Bi-LSTM. Langkah pertama dimulai dari pemuatan data historis pelunasan biaya haji reguler tahun 2024 yang diperoleh dari Kantor Kementerian Agama Kota Malang. Dataset ini memuat informasi penting yang dibutuhkan dalam proses prediksi, meliputi: Nomor Porsi, Jenis Kelamin, Umur, Pendidikan, Pekerjaan, Bulan Daftar, Tahun Daftar, Bulan Lunas, Tahun Lunas, Jangka Waktu Pelunasan, dan Status Pelunasan. Setelah data berhasil dimuat, dilakukan proses pembersihan data untuk menangani nilai kosong (*missing values*), data duplikat, atau nilai ekstrem yang tidak logis. Untuk fitur kategorik seperti jenis kelamin, pendidikan, dan pekerjaan, dilakukan proses encoding ke dalam format numerik. Hal ini karena model deep learning seperti Bi-LSTM hanya dapat memproses data numerik secara langsung. Dengan encoding, fitur kategorik dapat dikenali sebagai bagian dari input vektor dalam pembelajaran model.

Selanjutnya, data menjalani proses normalisasi dengan metode *Min-Max Scaling* untuk menyetarakan skala antar fitur numerik seperti umur dan jangka waktu pelunasan agar bernilai dalam rentang 0 hingga 1. Normalisasi ini diperlukan agar fitur yang memiliki skala besar tidak mendominasi pembelajaran model. Perhitungan normalisasi *min-max* dapat dinyatakan dalam persamaan 3.1 berikut :

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

Keterangan :

- a) x' : Nilai hasil normalisasi dengan rentang 0-1
- b) x : Nilai asli
- c) x_{min} : Nilai *minimum*
- d) x_{max} : Nilai *maximum*

Berikut disajikan contoh perhitungan manual normalisasi pada fitur “Umur” kolom pertama $x = 60$ pada Tabel 3.1 Sampel data nilai *minimum* (x_{min}) = 29 dan nilai *maximum* (x_{max}) = 62, maka perhitungan manual nilai normalisasi dari umur 60 adalah 0,9394:

$$x' = \frac{60 - 29}{62 - 29} = \frac{31}{33} \approx 0.9394$$

Tahap selanjutnya dalam penelitian ini adalah proses pembagian data (*data splitting*), yaitu memisahkan dataset menjadi data latih dan data uji untuk keperluan pelatihan dan pengujian model. Penelitian ini menerapkan dua rasio, yaitu 80:20 dan 70:30, kemudian menggunakan pendekatan *k-fold cross-validation* sebanyak 10 kali untuk masing-masing rasio pembagian data. Teknik ini bertujuan untuk menguji kestabilan dan konsistensi model secara lebih menyeluruh dengan membagi dataset menjadi sepuluh bagian, di mana setiap bagian secara bergantian digunakan sebagai data uji dan sisanya sebagai data latih. Penerapan *k-fold cross-validation* dinilai mampu memberikan hasil evaluasi yang lebih representatif terhadap kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

3.4 Membangun Model BI-LSTM

Model Bi-LSTM dibangun berdasarkan unit LSTM standar yang terdiri dari tiga gerbang utama: *forget gate*, *input gate*, dan *output gate*. Setiap *gate* memiliki fungsi tersendiri untuk mengatur aliran informasi di dalam memori jaringan. Proses ini memungkinkan model mempertahankan informasi penting dan mengabaikan informasi yang tidak relevan. Perhitungan untuk masing-masing gerbang dapat dijelaskan sebagai berikut:

a. *Forget Gate*

Forget gate berfungsi menentukan bagian informasi dari *cell state* sebelumnya yang perlu dilupakan. Gerbang ini menggunakan fungsi *sigmoid* untuk menghasilkan nilai antara 0 dan 1. Nilai 0 berarti informasi sepenuhnya dilupakan, sementara nilai 1 berarti informasi sepenuhnya dipertahankan. Rumusnya adalah sebagai berikut:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3.1)$$

Keterangan :

- a) f_t : nilai *forget gate*
- b) σ : fungsi *sigmoid*
- c) W_{fx} : bobot *forget gate* yang didapat selama pelatihan
- d) h_{t-1} : *hidden state* dari waktu sebelumnya
- e) x_t : input pada waktu t
- f) b_f : *forget gate bias*

b. *Input Gate*

Input gate bertugas menentukan informasi baru yang akan disimpan ke dalam *cell state*. Proses ini terdiri dari dua bagian: satu lapisan *sigmoid* untuk menentukan nilai yang akan diperbarui, dan satu lapisan *tanh* untuk membuat vektor kandidat nilai baru. Rumusnya adalah:

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.2)$$

$$\tilde{C}_t = \tanh (W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (3.4)$$

Keterangan :

- a) i_t : nilai *input gate*
- b) σ : fungsi *sigmoid*
- c) $\tilde{C}_t$: nilai *memory candidate*
- d) W_c : bobot memori kandidat

c. *Output Gate*

Output Gate berfungsi untuk memutuskan informasi dari *cell state* yang akan diteruskan sebagai output dari LSTM pada waktu saat ini. Gerbang ini bekerja dengan cara mengambil *cell state* yang diperbarui dan mengaplikasikan fungsi aktivasi *tanh* untuk menghasilkan representasi dari *cell state*. Hasil ini kemudian dikalikan dengan output dari *output gate*, yang diproses dengan fungsi aktivasi *sigmoid*. Rumus untuk *output gate* dapat dilihat pada persamaan 3.5 dan *hidden state* pada 3.6.

$$o_t = \sigma (W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (3.6)$$

Keterangan :

- a) o_t : nilai *output gate*
- b) σ : fungsi *sigmoid*
- c) h_t : nilai *hidden state*
- d) C_t : *cell state*

d. *Penggabungan Gate*

Bi-LSTM bekerja dengan menggabungkan dua lapisan LSTM yang berjalan secara paralel namun dengan arah yang berbeda. Lapisan pertama memproses data dari awal hingga akhir (*forward*), sedangkan lapisan kedua memproses data dari

akhir hingga awal (*backward*). Penggabungan ini dilakukan melalui operasi *concatenation*, di mana hasil dari kedua LSTM tersebut digabungkan menjadi satu vektor panjang. Penggabungan output dalam Bi-LSTM dapat dinyatakan dalam persamaan 3.7 .

$$y_t = (\overrightarrow{h_t}; \overleftarrow{h_t}) \quad (3.7)$$

$$\hat{y} = \sigma (W_d \cdot h_t^{(Bi-LSTM)} + b_d) \quad (3.8)$$

Keterangan :

- a) y_t : vektor gabungan
- b) $;$: fungsi *concatenation*
- c) $\overrightarrow{h_t}$: *hidden states* pada *forward pass*
- d) $\overleftarrow{h_t}$: *hidden states* pada *backward pass*
- e) W_d : Bobot *dense layer*
- f) $h_t^{(Bi-LSTM)}$: *Hidden state* dari Bi-LSTM

3.4.1 Prediksi dengan BI-LSTM

Setelah data melalui tahap *preprocessing*, model Bi-LSTM dilatih menggunakan dataset latih. Proses pelatihan memanfaatkan *Adam Optimizer* untuk mempercepat konvergensi dan menghindari local minima. Fungsi aktivasi *sigmoid* digunakan pada lapisan output karena target prediksi bersifat biner (melunasi atau tidak melunasi) (Kumar & Garg, 2021).

Untuk memperjelas mekanisme kerja Bi-LSTM, berikut disajikan contoh perhitungan dengan satu baris data yang telah melalui proses *encoding* dan normalisasi.

Tabel 3.2 Contoh data input yang telah dinormalisasi dan di*encoding*

Fitur	Nilai
Jenis Kelamin	1.0
Umur	0.538
Pendidikan	0.285
Pekerjaan	0.615
Bulan Daftar	0.0

Fitur	Nilai
Tahun Daftar	0.272
Bulan Lunas	0.333
Tahun Lunas	0.0
Jangka Waktu Pelunasan	0.7

Perhitungan dimulai dari menghitung *forget gate* pada persamaan (3.1) menggunakan input $x_t=[1.0,0.538,0.285,0.615,0.0,0.272,0.333,0.0,0.7]$ dengan inisialisasi bobot (W) dan bias (b) yaitu : $W_f= [0.2, 0.1, 0.3, 0.4, 0.2, 0.1, 0.3, 0.2, 0.1, 0.2, 0.3]$ dan $b_f = [0.1]$. Pada perhitungan ini inisialisasi *hidden state* sebelumnya $h_{t-1}= 0$ dan *cell state* sebelumnya $C_{t-1}= 0$, menggunakan hidden size = 1, maka diperoleh perhitunganya :

Forget gate :

$$ft = \sigma(Wf[ht-1, xt] + bf)$$

$$ft = \sigma(0.2, 0.1, 0.3, 0.4, 0.2, 0.1, 0.3, 0.2, 0.1, 0.2, 0.3$$

$$[0,1.0,0.538,0.285,0.615,0.0,0.272,0.333,0.0,0.7] + 0.1)$$

$$ft = \sigma((0+0.1+0.1614+0.114+0.123+0+0.0816+0.0666+0+0.14)+0.1)$$

$$ft = \sigma(0.7866+0.1)$$

$$ft = \sigma(0.8866)$$

$$ft = 0.7085$$

Selanjutnya menghitung nilai *input gate* pada persamaan (3.2) menggunakan $W_i=[0.3,0.2,0.2,0.1,0.4,0.3,0.2,0.1,0.2,0.1]$ dan $b_i = 0.05$, didapatkan hasil= 0.687 dengan perhitungan yang diperoleh :

Input gate :

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$i_t = \sigma (0.3, 0.2, 0.2, 0.1, 0.4, 0.3, 0.2, 0.1, 0.2, 0.1 \cdot$$

$$[0, 1.0, 0.538, 0.285, 0.615, 0.0, 0.272, 0.333, 0.0, 0.7] + 0.05)$$

$$i_t = \sigma ((0 + 0.2 + 0.1076 + 0.0285 + 0.246 + 0 + 0.0544 + 0.0333 + 0 + 0.07) + 0.05)$$

$$i_t = \sigma (0.7398 + 0.05)$$

$$i_t = \sigma (0.7898)$$

$$i_t = 0.687$$

Setelah melakukan perhitungan *input gate*, selanjutnya menghitung kandidat *cell state*. Perhitungan kandidat *cell state* ditunjukkan pada persamaan (3.3) dengan menggunakan nilai $W_c = [0.2, 0.3, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3, 0.1]$ dan $b_c = 0.05$, mendapat nilai sebesar 0.610.

Kandidat *CellState* :

$$C_t^{\sim} = \tanh (W_c \cdot [h_{t-1} \cdot x_t] + b_c)$$

$$C_t^{\sim} = \tanh ((0.2, 0.3, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3, 0.1) \cdot$$

$$[0, 1.0, 0.538, 0.285, 0.615, 0.0, 0.272, 0.333, 0.0, 0.7] + 0.05)$$

$$C_t^{\sim} = \tanh ((0 + 0.3 + 0.0538 + 0.0855 + 0.0615 + 0 + 0.0544 + 0.0333 + 0 + 0.07) + 0.05)$$

$$C_t^{\sim} = \tanh (0.6585 + 0.05)$$

$$C_t^{\sim} = \tanh (0.7085)$$

$$C_t^{\sim} = 0.610$$

Tahap selanjutnya yaitu perhitungan *cell state* yang ditunjukkan pada persamaan (3.4) menggunakan nilai *forget gate*, *input gate* dan kandidat *cell state* yang sudah dihitung sebelumnya.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t^{\sim}$$

$$C_t = (0.7085 \cdot 0) + (0.687 \cdot 0.610)$$

$$C_t = 0.41907$$

Pada langkah ini untuk menghitung nilai *output gate* pada persamaan (3.5) dengan menggunakan $W_o = [0.3, 0.2, 0.1, 0.2, 0.1, 0.3, 0.2, 0.2, 0.1, 0.3]$, $b_o = 0.1$ dan $h_{t-1}, x_t = [0, 1.0, 0.538, 0.285, 0.615, 0.0, 0.272, 0.333, 0.0, 0.7]$ sebagai berikut :

Output gate :

$$o_t = \sigma (W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$o_t = \sigma (0.3, 0.2, 0.1, 0.2, 0.1, 0.3, 0.2, 0.2, 0.1, 0.3) \cdot [0, 1.0, 0.538, 0.285, 0.615, 0.0, 0.272, 0.333, 0.0, 0.7] + 0.1)$$

$$o_t = \sigma ((0 + 0.2 + 0.0538 + 0.057 + 0.0615 + 0 + 0.0544 + 0.0666 + 0 + 0.21) + 0.1)$$

$$o_t = \sigma (0.7033 + 0.1)$$

$$o_t = \sigma (0.8033)$$

$$o_t = 0.691$$

Lalu yang terakhir adalah menghitung *hidden state* dengan hasil *output gate* yang sudah dihitung dikalikan dengan hasil *cell state* yakni $o_t = 0.691$ dan $C_t = 0.41907$ Perhitungan *hidden state* ditunjukkan pada persamaan (3.5).

Hidden state :

$$h_1 = o_t \cdot \tanh(C_t)$$

$$h_1 = 0.691 \cdot \tanh(0.41907)$$

$$h_1 = 0.691 \cdot 0.396$$

$$h_1 = 0.274$$

Sudah didapatkan hasil nilai *hidden state* (h_1) pada perhitungan LSTM *forward*, kemudian kita lanjutkan dengan menghitung *forget gate*, *input gate*, *output gate backward*.

Perhitungan dimulai dari menghitung *forget gate* pada persamaan (3.1) menggunakan input dalam bentuk urutan yang dibalik (reverse) $x_t^{(backward)}$ $= [0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0]$ dengan inisialisasi bobot (W) dan bias (b) yaitu : $W_f = [0.2, 0.1, 0.3, 0.4, 0.2, 0.1, 0.3, 0.2, 0.1, 0.2, 0.3]$ dan $b_f = [0.1]$. Pada perhitungan ini inisialisasi *hidden state* sebelumnya $h_{t-1} = 0$ dan *cell state* sebelumnya $C_{t-1} = 0$, menggunakan hidden size = 1, maka diperoleh perhitunganya sebagai berikut :

Forget gate backward :

$$f_t^{(backward)} = \sigma(Wf[ht-1, xt] + bf)$$

$$f_t^{(backward)} = \sigma(0.2, 0.1, 0.3, 0.4, 0.2, 0.1, 0.3, 0.2, 0.1, 0.2, 0.3 \\ [0, 0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0] + 0.1)$$

$$f_t^{(backward)} =$$

$$\sigma((0+0.07+0+0.1332+0.0544+0+0.1845+0.057+0.0538+0.2)+0.1)$$

$$f_t^{(backward)} = \sigma(0.7529+0.1)$$

$$f_t^{(backward)} = \sigma(0.8529)$$

$$f_t^{(backward)} = 0.701$$

Selanjutnya menghitung nilai *input gate backward* pada persamaan (3.2) menggunakan $W_i = [0.3, 0.2, 0.2, 0.1, 0.4, 0.3, 0.2, 0.1, 0.2, 0.1]$ dan $b_i = 0.05$, didapatkan hasil = 0.687 dengan perhitungan yang diperoleh :

Input gate backward :

$$i_t^{(backward)} = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$i_t^{(backward)} = \sigma (0.3, 0.2, 0.2, 0.1, 0.4, 0.3, 0.2, 0.1, 0.2, 0.1 \cdot$$

$$[0, 0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0] + 0.05)$$

$$i_t^{(backward)} = \sigma ((0 + 0.14 + 0 + 0.0333 + 0.1088 + 0 + 0.123 + 0.0285 + 0.1076 + 0.1) + 0.05)$$

$$i_t^{(backward)} = \sigma (0.6412 + 0.05)$$

$$i_t^{(backward)} = \sigma (0.6912)$$

$$i_t^{(backward)} = 0.666$$

Setelah melakukan perhitungan *input gate backward*, selanjutnya menghitung kandidat *cell state*. Perhitungan kandidat *cell state* ditunjukkan pada persamaan (3.3) dengan menggunakan nilai $W_c = [0.2, 0.3, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3, 0.1]$ dan $b_c = 0.05$, mendapat nilai sebesar 0.610.

Kandidat *CellState* :

$$C_t^{\sim (backward)} = \tanh (W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$C_t^{\sim (backward)} = \tanh ((0.2, 0.3, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3, 0.1) \cdot$$

$$[0, 0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0] + 0.05)$$

$$C_t^{\sim (backward)} = \tanh ((+0.21 + 0 + 0.0999 + 0.0272 + 0 + 0.123 + 0.0285 + 0.1614 + 0.1) + 0.05)$$

$$C_t^{\sim (backward)} = \tanh (0.749 + 0.05)$$

$$C_t^{(backward)} = \tanh(0.799)$$

Tahap selanjutnya yaitu perhitungan *cell state* yang ditunjukkan pada persamaan (3.4) menggunakan nilai $f_t^{(backward)} = 0.701$, $C_{t-1} = 0$, $i_t^{(backward)} = 0.666$, dan $C_t^{(backward)} = 0.663$ yang sudah dihitung sebelumnya.

Cell State backward:

$$C_t^{(backward)} = f_t \cdot C_{t-1} + i_t \cdot C_t^{\sim}$$

$$C_t^{(backward)} = (0.701 \cdot 0) + (0.666 \cdot 0.663)$$

$$C_t^{(backward)} = 0.441$$

Pada langkah ini untuk menghitung nilai *output gate backward* pada persamaan (3.5) dengan menggunakan $W_o = [0.3, 0.2, 0.1, 0.2, 0.1, 0.3, 0.2, 0.2, 0.1, 0.3]$, $b_o = 0.1$ dan $h_{t-1}, x_t = [0, 0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0]$ sebagai berikut :

Output gate backward:

$$o_t^{(backward)} = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$o_t^{(backward)} = \sigma((0.3, 0.2, 0.1, 0.2, 0.1, 0.3, 0.2, 0.2, 0.1, 0.3 \cdot$$

$$[0, 0.7, 0.0, 0.333, 0.272, 0.0, 0.615, 0.285, 0.538, 1.0] + 0.1)$$

$$o_t^{(backward)} = \sigma((0 + 0.14 + 0 + 0.0666 + 0.0272 + 0 + 0.123 + 0.057 + 0.0538 + 0.3) + 0.1)$$

$$o_t^{(backward)} = \sigma(0.7676 + 0.1)$$

$$o_t^{(backward)} = \sigma(0.8676)$$

$$o_t^{(backward)} = 0.704$$

Lalu yang terakhir adalah menghitung *hidden state backward* dengan hasil *output gate backward* yang sudah dihitung dikalikan dengan hasil *cell state backward* yakni $o_t^{(backward)}=0.704$ dan $C_t^{(backward)} = 0.441$. Perhitungan *hidden state backward* ditunjukkan pada persamaan (3.5).

Hidden state backward :

$$h_1^{(backward)} = o_t \cdot \tanh(C_t)$$

$$h_1^{(backward)} = 0.704 \cdot \tanh(0.441)$$

$$h_1^{(backward)} = 0.704 \cdot 0.414$$

$$h_1^{(backward)} = 0.291$$

Selanjutnya, untuk menghasilkan representasi akhir dari Bi-LSTM pada waktu y_t , nilai *hidden state* dari arah forward dan backward digabungkan. Proses penggabungan ini ditunjukkan pada Persamaan (3.6), dengan cara menggabungkan secara *concatenate* antara $\vec{h}_t$ dan $\overleftarrow{h}_t$. Adapun hasil akhir dari gabungan *hidden state* Bi-LSTM adalah:

$$y_t = (\vec{h}_t; \overleftarrow{h}_t)$$

$$h_t^{(Bi-LSTM)} = (0.274, 0.291)$$

Setelah digabungkan kemudian akan dihitung pada *dense layer* yang nantinya akan menghasilkan prediksi 1 = melunasi dan 0 = tidak melunasi. Perhitungan dense layer ditunjukkan pada persamaan (3.7) dengan menggunakan $W_d=[0.5, 0.4]$, dan $b_d= -0.2$ berikut diperoleh perhitungannya :

$$\hat{y} = \sigma (W_d \cdot h_t^{(Bi-LSTM)} + b_d)$$

$$\hat{y} = \sigma ((0.5, 0.4) \cdot (0.274, 0.291) + -0.2)$$

$$\hat{y} = \sigma ((0.5 \times 0.274) + (0.4 \times 0.291)) + -0.2)$$

$$\hat{y} = \sigma ((0.137 + 0.1164) + -0.2)$$

$$\hat{y} = \sigma (0.2534 + -0.2)$$

$$\hat{y} = \sigma (0.0534)$$

$$\hat{y} = 0.512$$

Fungsi aktivasi yang digunakan pada lapisan ini adalah fungsi *sigmoid*, yang secara matematis dirumuskan sebagai:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.9)$$

Fungsi sigmoid akan mengubah input menjadi nilai probabilitas antara 0 dan 1. Dalam konteks klasifikasi biner, jika nilai output $\hat{y} \geq 0.5$, maka model memprediksi kelas 1. Sebaliknya, jika $\hat{y} \leq 0.5$, maka diprediksi sebagai kelas 0 (Tollner et al., 2024).

Pada penelitian ini, fungsi *sigmoid* digunakan di lapisan *dense* terakhir untuk menghasilkan nilai prediksi akhir yakni 1 (melunasi) dan 0 (tidak melunasi) berdasarkan hasil dari penggabungan *hidden state* arah *forward* dan *backward* dari arsitektur Bi-LSTM. Pada perhitungan yang telah dilakukan sebelumnya, diperoleh nilai prediksi akhir sebesar 0,512. Karena nilai ini lebih besar dari ambang batas 0,5, maka model memprediksi bahwa calon jamaah tersebut akan melunasi.

3.5 Evaluasi Model

Evaluasi model merupakan langkah penting dalam proses pengembangan model machine learning, termasuk model Bi-LSTM. Tujuan dari evaluasi ini adalah untuk mengukur sejauh mana model dapat memprediksi hasil yang sesuai dengan data aktual. Proses evaluasi dilakukan setelah model selesai dilatih menggunakan data pelunasan jamaah haji, dengan membandingkan hasil prediksi model terhadap label aktual pada data validasi.

3.5.1 *Confusion Matrix*

Confusion Matrix adalah metode evaluasi untuk performa model machine learning, termasuk model Bi-LSTM dalam mencari masalah prediksi dengan keluaran berupa dua kelas atau lebih. *Confusion Matrix* menggambarkan hasil prediksi model dalam bentuk tabel, dimana setiap baris mewakili kelas sebenarnya, dan setiap kolom mewakili prediksi model (Phillips et al., 2024).

Tabel 3.3 *Confusion Matrix*

		True Values	
		True	False
Prediction	True	TP Correct Result	FP Unexpexted Result
	False	FN Missing Result	TN Correct Absence of Result

Ada empat komponen utama yang merupakan representasi hasil prediksi pada *confusion matrix* :

- True Positive* (TP): mendefinisikan jamaah haji yang melunasi (positif) dan diklasifikasikan sebagai melunasi (positif).
- False Positive* (FP): mendefinisikan jamaah haji yang tidak melunasi (negatif) diklasifikasikan sebagai melunasi (melunasi).

- c. *True Negative* (TN): mendefinisikan jamaah haji yang tidak melunasi (negatif) dan diklasifikasikan tidak melunasi (negatif).
- d. *False Negative* (FN): mendefinisikan jamaah haji melunasi (positif) dan diklasifikasikan tidak melunasi (negatif).

Dari hasil *confusion matrix*, diperoleh empat matrix utama yang dihitung tiap nilainya menggunakan :

1. **Accuracy**: Menggambarkan seberapa tepat model dalam memprediksi benar dibandingkan dengan seluruh data yang diuji.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3.10)$$

2. **Precision**: Menggambarkan tingkat ketepatan prediksi positif benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model.

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (3.11)$$

3. **Recall** : Menggambarkan seberapa baik model dalam menemukan seluruh data yang benar-benar positif.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (3.12)$$

4. **F1-score**: Menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan.

$$\text{F1 - Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \quad (3.13)$$

3.5.2 *K-fold Cross Validation*

K-fold Cross Validation dilakukan dengan cara memisahkan data menjadi dua bagian yakni data latih dan data validasi. Dalam teknik ini, *dataset* akan dilakukan pembagian ke dalam K bagian (*folds*) dengan ukuran yang sama. Setiap K partisi akan dilatih menggunakan K-1 data dan sisanya sebagai data validasi. Dalam penelitian ini penelitian dilakukan dengan menggunakan *K-fold* 10 (Dewi & Putra, 2024). Gambar 2.3 adalah gambar skema partisi data pada *K-Fold Cross Validation* dengan menggunakan *K-fold* 10.

KFold	Cross Validation									
1	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
2	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
3	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
4	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
5	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
6	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
7	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
8	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
9	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Gambar 3.3 Skema partisi data dengan K = 10

Keunggulan *K-Fold Cross Validation* adalah kemampuannya dalam mengurangi bias evaluasi yang mungkin timbul ketika hanya menggunakan satu subset data untuk pengujian (Endang Etriyanti, 2021).

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan hasil ujicoba, pembahasan, dan prediksi pelunasan haji di kemenag kota malang menggunakan metode *Bidirectional Long Short-Term Memory*

4.1 Data Input

Langkah pertama sebelum melakukan pengujian adalah mempersiapkan data yang akan digunakan. Sebagaimana yang telah disebutkan dalam Bab 3, data yang dimanfaatkan pada penelitian ini berasal dari data calon jamaah haji yang telah melakukan pelunasan haji, data diperoleh melalui Kemenag Kota Malang. Dataset ini terdiri dari 1.619 data dengan 11 fitur yang dijadikan sebagai label.

Tabel 4.1 Data pelunasan biaya haji 2023-2024

Porsi	Jenis Kelamin	Umur	Pendidikan	Pekerjaan	Bulan Daftar	Tahun Daftar	Bulan Lunas	Tahun Lunas	Jangka Waktu Pelunasan	Status Pelunasan
1300502734	L	60	S1	Pegawai Negri Sipil	1	2011	2	2023	12	Melunasi
1300590684	L	56	S1	Swasta	12	2011	1	2023	12	Melunasi
1300590686	P	55	S1	Swasta	12	2011	1	2023	12	Melunasi
1300575727	P	62	SLTA	Ibu Rumah Tangga	12	2011	1	2023	12	Tidak Melunasi
1300590920	L	29	SD	Swasta	12	2011	2	2023	12	Melunasi

4.1.1 Pembersihan Data

Pembersihan data adalah langkah awal yang sangat penting dalam proses analisis data (Sudarsono et al., 2021). Pada tahap ini, dilakukan berbagai pengecekan untuk memastikan data yang digunakan tidak mengandung nilai yang hilang (*missing values*) atau data yang tidak *valid* yang dapat mengganggu analisis selanjutnya. *Missing values* sering kali terjadi akibat kesalahan dalam proses input data atau ketidaksesuaian saat pengumpulan data (Ningrum et al., 2021). Oleh karena itu, penting untuk melakukan pengecekan terhadap nilai-nilai yang hilang ini agar data yang digunakan lebih *valid* dan akurat.

Tabel 4.2 Pengecekan nilai yang hilang

No	Kolom	Jumlah <i>Missing</i>
0	Porsi	0
1	Jenis Kelamin	0
2	Umur	0
3	Pendidikan	0
4	Pekerjaan	0
5	Bulan Daftar	0
6	Tahun Daftar	0
7	Bulan Lunas	0
8	Tahun Lunas	0
9	Jangka Waktu Pelunasan	0
10	Status Pelunasan	0

Pengecekan *missing values* ini dapat dilihat pada Tabel 4.2, yang menunjukkan jumlah nilai yang hilang pada setiap kolom dalam dataset. Hasilnya memberikan informasi mengenai kolom mana saja yang mengandung nilai yang hilang serta jumlahnya, yang dapat digunakan untuk langkah penanganan lebih lanjut. Selain itu, dilakukan pengecekan terhadap tipe data pada setiap kolom untuk memastikan bahwa tipe data yang digunakan sesuai dengan jenis data yang diinginkan, seperti kolom numerik yang berisi angka dan kolom kategorikal yang

berisi kategori. Pengecekan ini bertujuan untuk menghindari kesalahan dalam pengolahan data lebih lanjut.

Tabel 4.3 Hasil pengecekan tipe data

No	Kolom	Tipe Data
0	Porsi	<i>int64</i>
1	Jenis Kelamin	<i>string</i>
2	Umur	<i>int64</i>
3	Pendidikan	<i>string</i>
4	Pekerjaan	<i>string</i>
5	Bulan Daftar	<i>int64</i>
6	Tahun Daftar	<i>int64</i>
7	Bulan Lunas	<i>int64</i>
8	Tahun Lunas	<i>int64</i>
9	Jangka Waktu Pelunasan	<i>int64</i>
10	Status Pelunasan	<i>string</i>

Selain itu, pengecekan terhadap format data dilakukan untuk memastikan bahwa setiap kolom dalam dataset memiliki tipe data yang tepat dan sesuai dengan yang diinginkan. Pengecekan ini bertujuan untuk menghindari adanya ketidaksesuaian tipe data yang bisa menyebabkan masalah saat data diproses dalam model analisis. Hasil pengecekan tipe data disajikan dalam Tabel 4.3, yang memberikan gambaran mengenai tipe data yang terdapat pada setiap kolom dalam dataset.

Proses *encoding* seperti yang dapat dilihat pada Tabel 4.4 dilakukan untuk mengonversi data kategorikal ke dalam bentuk numerik agar dapat diolah oleh algoritma machine learning. Pada atribut Jenis Kelamin, dilakukan konversi nilai 'L' yang merepresentasikan laki-laki menjadi 1, dan 'P' yang merepresentasikan perempuan menjadi 0. Selanjutnya, pada atribut Pendidikan, jenjang pendidikan dikategorikan secara berurutan berdasarkan tingkatannya. Pendidikan dasar seperti 'SD' diberi nilai 0, kemudian 'SLTP' menjadi 1, 'SLTA' menjadi 2, sementara

kategori seperti 'D1 / D2 / D3 / SM' serta 'D3' disamakan menjadi 3. Untuk jenjang pendidikan yang lebih tinggi, 'S1' diberi nilai 4, 'S2' menjadi 5, dan 'S3' diberi nilai 6. Adapun pada atribut Pekerjaan, berbagai jenis pekerjaan juga dikonversi ke dalam nilai numerik. 'Pegawai Negeri Sipil' dikodekan sebagai 0, 'Swasta' sebagai 1, 'Ibu Rumah Tangga' sebagai 2, 'Dagang' sebagai 3, 'TNI / Polri' sebagai 4, 'Tani' sebagai 5, 'Pensiunan' sebagai 6, 'BUMN' sebagai 7, 'Lainnya' sebagai 8, 'Mahasiswa' sebagai 9, dan 'Nelayan' sebagai 10.

Tabel 4.4 Hasil proses *encoding*

Porsi	Jenis Kelamin	Umur	Pendidikan	Pekerjaan	Bulan Daftar	Tahun Daftar	Bulan Lunas	Tahun Lunas	Jangka Waktu Pelunasan	Status Pelunasan
1300502734	1	60	0	4	1	2011	2	2023	12	1
1300590684	1	53	1	4	5	2011	2	2023	12	1
1300590686	0	68	1	4	9	2011	1	2023	12	1
1300575727	0	69	2	2	9	2011	1	2023	12	0
1300590920	1	53	1	0	10	2011	1	2023	12	1

4.1.2 Pemilihan Fitur

Pemilihan fitur merupakan tahap penting dalam proses pemodelan data, khususnya pada sistem pembelajaran mesin, karena fitur yang dipilih akan menentukan informasi apa saja yang akan dipelajari oleh model untuk menghasilkan prediksi. Dalam konteks penelitian ini, pemilihan fitur bertujuan untuk menyaring dan mengelompokkan data menjadi dua bagian utama, yaitu fitur – fitur yang digunakan untuk prediksi (x) dan label target (y). Proses ini dilakukan agar struktur data sesuai dengan kebutuhan *input* dan *output* pada model Bi-LSTM yang digunakan dalam penelitian.

Tabel 4.5 Fitur - fitur yang dipilih untuk prediksi

Jenis Kelamin	Umur	Pekerjaan	Pendidikan	Bulan Daftar	Tahun Daftar	Bulan Lunas	Tahun Lunas	Jangka Waktu Pelunasan
1	60	4	0	1	2011	2	2023	12
1	53	4	1	5	2011	2	2023	12
0	68	4	1	9	2011	1	2023	12
0	69	2	2	9	2011	1	2023	12
1	53	0	1	10	2011	1	2023	12

Pada Tabel 4.5 diatas Kesembilan fitur dipilih menjadi parameter karena mampu merepresentasikan karakteristik calon jemaah serta proses pelunasan biaya haji secara menyeluruh. Misalnya, variabel seperti pendidikan dan pekerjaan dapat memberikan gambaran tentang latar belakang sosial dan ekonomi calon jemaah. Sementara itu, fitur waktu seperti bulan dan tahun daftar atau lunas, serta jangka waktu pelunasan, berguna untuk melihat pola keterlambatan atau kecepatan dalam melunasi biaya haji. Dengan menyediakan informasi yang bersifat demografis dan temporal, model memiliki cukup data untuk mengenali pola-pola yang relevan dalam proses pelunasan biaya haji.

Sementara itu, variabel y dalam penelitian ini adalah Status Pelunasan, yang merepresentasikan label biner dari data, yaitu apakah calon jemaah telah melunasi atau belum melunasi biaya hajinya. Variabel ini menjadi acuan bagi model untuk melakukan klasifikasi dan menjadi target utama yang ingin diprediksi melalui serangkaian proses pembelajaran.

4.1.3 Normalisasi Data

Normalisasi data adalah proses penyesuaian skala nilai dari masing-masing fitur numerik agar berada dalam rentang tertentu, sehingga tidak ada fitur yang mendominasi yang lain hanya karena memiliki skala yang lebih besar. Proses ini sangat penting terutama dalam penerapan algoritma pembelajaran mesin yang berbasis jaringan saraf seperti Bi-LSTM, karena perbedaan skala antar fitur dapat mempengaruhi kestabilan dan performa model.

Tabel 4.6 Nilai minimum dan maksimum setiap fitur sebelum Normalisasi

No	Fitur	Nilai Minimum	Nilai Maksimum
0	Jenis Kelamin	0	1
1	Umur	18	95
2	Pendidikan	0	6

3	Pekerjaan	0	10
4	Bulan Daftar	1	12
5	Tahun Daftar	2008	2019
6	Bulan Lunas	1	4
7	Tahun Lunas	2023	2024
8	Jangka Waktu Pelunasan	5	15

Dalam penelitian ini, normalisasi dilakukan hanya terhadap fitur-fitur numerik dapat dilihat pada Tabel 4.6, sedangkan label target (y) tidak melalui proses normalisasi karena bersifat kategorikal biner. Metode yang digunakan untuk melakukan normalisasi adalah *Min-Max Scaling*, yang merupakan salah satu teknik dari *Scikit-learn*. *MinMaxScaler* bekerja dengan memetakan nilai terkecil dari fitur menjadi 0 dan nilai terbesar menjadi 1, sehingga semua nilai fitur akan berada dalam rentang [0, 1].

Tabel 4.7 Hasil Normalisasi data

No	Jenis Kelamin	Umur	Pendidikan	Pekerjaan	Bulan Daftar	Tahun Daftar	Bulan Lunas	Tahun Lunas	Jangka Waktu Pelunasan
0	1.0	0.545455	0.666667	0.0	0.000000	0.272727	0.333333	0.0	0.7
1	1.0	0.493506	0.666667	0.1	1.000000	0.272727	0.000000	0.0	0.7
2	0.0	0.480519	0.666667	0.1	1.000000	0.272727	0.000000	0.0	0.7
3	0.0	0.571429	0.333333	0.2	1.000000	0.272727	0.000000	0	0.7
4	1.0	0.142857	0.000000	0.1	1.000000	0.272727	0.333333	0.0	0.7
5	0.0	0.363636	0.333333	0.0	0.181818	0.363636	1.000000	1.0	0.7
6	1.0	0.610390	0.000000	0.3	0.090909	0.363636	0.333333	0.0	0.6
7	1.0	0.610390	0.833333	0.1	0.090909	0.363636	0.333333	0.0	0.6
8	0.0	0.506494	0.333333	0.2	0.090909	0.363636	0.333333	0.0	0.6
9	0.0	0.623377	0.166667	0.2	0.090909	0.363636	0.333333	0.0	0.6

Pada Tabel 4.7 menampilkan secara rinci daftar fitur numerik yang dinormalisasi dalam penelitian ini menggunakan metode *Min-Max Scaling*. Fitur-fitur numerik yang dinormalisasi dalam penelitian ini antara lain: “Umur”, “Bulan Daftar”, “Tahun Daftar”, “Bulan Lunas”, “Tahun Lunas”, dan “Jangka Waktu Pelunasan”. Sedangkan fitur kategorikal seperti “Jenis Kelamin”, “Pendidikan”, dan “Pekerjaan” terlebih dahulu dikonversi menjadi angka melalui proses *Label Encoding*.

4.2 Pemodelan dan Pelatihan BI-LSTM

Proses pelatihan model pada penelitian ini menggunakan pembagian rasi data yang ditetapkan untuk mengevaluasi performa model. Proses pelatihan menggunakan split data dengan dua rasio yaitu 80:20 (80% data *training* dan 20% data *testing*), dan 70:30 (70% data *training* dan 30% data *testing*) kemudian menggunakan *k-fold cross validation* dengan K yaitu 10. Pengujian ini dilakukan bertujuan untuk menentukan evaluasi model untuk mencapai akurasi yang terbaik dalam memprediksi pelunasan biaya haji dengan data yang digunakan diambil dari KEMENAG (Kementerian Agama) Kota Malang.

Tabel 4.8 Pembagian data

Rasio	Data Training	Data Testing
70% : 30%	1.133	486
80% : 20%	1.295	324

4.3 Pengujian BI-LSTM

Pengujian model Bi-LSTM dilakukan untuk mengetahui sejauh mana kemampuan model dalam melakukan klasifikasi status pelunasan biaya haji

berdasarkan data yang telah melalui proses *preprocessing* dan *windowing*. Model Bi-LSTM ini dirancang untuk mengenali pola-pola sekuensial dalam data historis, sehingga dapat membuat prediksi status pelunasan secara lebih akurat.

Proses prediksi oleh model Bi-LSTM dimulai ketika model menerima input data dalam bentuk sekuens berdimensi tiga, yaitu jumlah *window*, panjang *window*, jumlah fitur. Setiap *window* berisi urutan nilai-nilai dari sembilan fitur yang telah dipilih, yang menggambarkan kondisi calon jemaah dalam rentang waktu tertentu. Fitur-fitur ini, seperti umur, pekerjaan, hingga bulan dan tahun pelunasan, dimasukkan sebagai representasi numerik agar dapat diproses oleh jaringan saraf.

Setiap unit LSTM di dalam jaringan memiliki memori internal yang mempertahankan informasi penting dari langkah-langkah sebelumnya dan mekanisme gerbang (*gates*) yakni *input gate*, *forget gate*, dan *output gate*, yang berfungsi mengatur aliran informasi. Dengan bantuan mekanisme LSTM, model dapat “mengingat” atau “melupakan” informasi sesuai kebutuhan, dan menghindari masalah umum seperti vanishing gradient pada jaringan RNN biasa.

Setelah sekuens selesai diproses oleh lapisan-lapisan Bi-LSTM, informasi terakhir yang dirangkum diteruskan ke lapisan *dense*, yang berfungsi melakukan transformasi akhir terhadap representasi internal model. Lapisan terakhir menggunakan fungsi aktivasi *sigmoid*, sehingga hasil output berupa nilai antara 0 dan 1, yang mewakili probabilitas status pelunasan. Jika nilai output mendekati 1, maka model mengklasifikasikan bahwa calon jemaah sudah melunasi, sedangkan jika mendekati 0, maka dianggap belum melunasi. Keputusan klasifikasi ini

kemudian dibandingkan dengan data aktual untuk menghitung metrik performa seperti *precision*, *recall*, dan *F1-score*.

4.4 Skenario Pengujian

Penelitian ini menerapkan pengujian berdasarkan pembagian data latih dan data uji (*train-test split*) dengan dua rasio, yaitu 80:20 dan 70:30. Pada rasio 80:20, sebanyak 80% data digunakan untuk proses pelatihan model, sementara 20% sisanya digunakan untuk pengujian. Demikian pula pada rasio 70:30, sebanyak 70% data digunakan sebagai data latih dan 30% sisanya sebagai data uji. Setelah pembagian data ini dilakukan, masing-masing subset data latih kemudian diuji menggunakan pendekatan *k-fold cross-validation* sebanyak 10 kali (*10-fold*). Penerapan teknik *k-fold cross-validation* pada data latih bertujuan untuk meningkatkan validitas evaluasi model dengan membagi data secara lebih merata dalam proses pelatihan dan pengujian, serta menghindari bias yang mungkin terjadi akibat pembagian data yang statis. Teknik ini memungkinkan model untuk dilatih dan diuji secara bergantian pada seluruh bagian subset data latih, sehingga hasil evaluasi yang diperoleh lebih representatif terhadap kinerja model secara keseluruhan.

Pada kedua pembagian data, model Bi-LSTM dibangun dengan parameter pelatihan yang sama, yaitu jumlah *epoch* sebanyak 50, *batch size* sebesar 16, *optimizer Adam*, dan fungsi *loss binary cross-entropy*. Model dilatih serta disimpan berdasarkan nilai *validation loss* terbaik menggunakan model *check pointing*. Evaluasi performa dilakukan menggunakan empat metrik utama, yaitu *precision*, *recall*, dan *F1-score*. Keempat metrik ini digunakan karena mampu memberikan

penilaian menyeluruh terhadap kualitas klasifikasi model, khususnya dalam konteks prediksi biner seperti status pelunasan (melunasi dan tidak melunasi). Selain itu, dilakukan juga visualisasi berupa *confusion matrix* untuk setiap skenario pengujian.

4.5 Hasil

Pengujian dilakukan dengan membagi data ke dalam beberapa rasio, yaitu 80:20 dan 70:30, serta menerapkan metode *k-fold cross-validation* sebanyak 10 kali guna memastikan model diuji secara menyeluruh dan adil di berbagai subset data. Setiap hasil pengujian dievaluasi berdasarkan metrik-metrik penting seperti *training loss*, *validation accuracy*, *precision*, *recall*, *F1-score*, dan akurasi keseluruhan untuk menilai kinerja model dalam mendeteksi kelas positif maupun negatif. Hasil pengujian ini menjadi dasar untuk menganalisis keunggulan serta keterbatasan model dalam memprediksi pelunasan biaya haji pada skenario yang berbeda.

4.5.1 Hasil Pengujian Rasio 80:20

Pada pembagian data 80:20, model dilatih dengan menggunakan *k-fold cross-validation* sebanyak 10 kali. Hasil yang diperoleh dapat dilihat pada Tabel 4.8 menunjukkan bahwa nilai *training loss* menurun di setiap *epoch*, menandakan proses pembelajaran model berjalan baik. Sementara itu, akurasi pengujian (*Validation accuracy*) cenderung meningkat seiring bertambahnya *epoch*, dengan performa terbaik umumnya dicapai pada *epoch* ke-5. Akurasi tertinggi tercatat pada

fold ke-10 sebesar 0.8516, sedangkan akurasi terendah berada pada *fold* ke-3 pada *epoch* kedua dengan nilai 0.5504.

Tabel 4.9 Hasil *Train. Loss* dan *Val. Accuracy* 80:20 dengan 10-Fold

<i>n-fold</i>	<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Accuracy</i>
1	1	1.4233	0.6977
	2	1.0785	0.6977
	3	0.8976	0.7364
	4	0.7724	0.7442
	5	0.7079	0.7829
2	1	1.4293	0.7209
	2	1.0647	0.7209
	3	0.8665	0.7752
	4	0.7534	0.7907
	5	0.6890	0.7907
3	1	1.3899	0.5504
	2	1.0250	0.7209
	3	0.8284	0.7907
	4	0.7274	0.7907
	5	0.6695	0.8062
4	1	1.4054	0.6124
	2	1.0610	0.6357
	3	0.8577	0.7054
	4	0.7413	0.7287
	5	0.6788	0.7674
5	1	1.3983	0.5581
	2	1.0206	0.6047
	3	0.8110	0.6667
	4	0.6972	0.6822
	5	0.6391	0.7442
6	1	1.4059	0.6512
	2	1.0695	0.6977
	3	0.8837	0.7287
	4	0.7883	0.7883
	5	0.7192	0.7519
7	1	1.4096	0.6875
	2	1.0282	0.6953
	3	0.8170	0.7656
	4	0.7175	0.7734
	5	0.6490	0.8125
8	1	1.3892	0.6406
	2	1.0016	0.6484
	3	0.8225	0.6719
	4	0.7118	0.7422
	5	0.6501	0.7734
9	1	1.3937	0.6250
	2	0.9820	0.6641
	3	0.8026	0.7422
	4	0.7056	0.7266
	5	0.6531	0.8125
10	1	1.3795	0.6484

<i>n-fold</i>	<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Accuracy</i>
	2	1.0231	0.7266
	3	0.8201	0.7969
	4	0.7083	0.8047
	5	0.6548	0.8516

Hasil rasio pengujian 80:20 dengan menggunakan 10-fold cross validation mendapatkan hasil perhitungan seperti pada Tabel 4.9. Evaluasi dilakukan dengan mengukur metrik klasifikasi seperti *precision*, *recall*, dan *F1-score* untuk masing-masing kelas positif dan negatif, serta nilai akurasi keseluruhan. Hasil menunjukkan bahwa nilai *precision* kelas positif berkisar antara 60,78% hingga 79,25%, sedangkan *precision* negatif antara 74,70% hingga 84,62%. Nilai *recall* untuk kelas positif mencapai titik tertinggi sebesar 80,77% pada *fold* ke-2, sementara *recall* negatif tertinggi sebesar 84,72% pada *fold* ke-7. *F1-score* terbaik untuk kedua kelas juga muncul pada *fold* ke-7, yakni 77,06% (positif) dan 82,99% (negatif), menunjukkan performa model paling seimbang pada *fold* tersebut.

Akurasi keseluruhan berada dalam rentang 70,50% hingga 80,47%, dengan rata-rata akurasi sebesar 75,20%. *Fold* terbaik pada rasio 80:20 berada pada *fold* ke 7 dengan nilai *precision* positif 79,25% , *precision* negatif 81,33%, *recall* positif 75%, *recall* negatif 84,72%, *F1-score* positif 77,06% , *F1-score* negatif 82,99% dan akurasi 80,47%. *Precision* positif 79,25% menunjukkan bahwa model cukup akurat dalam memprediksi kelas positif dengan sedikit kesalahan, meskipun masih ada beberapa prediksi positif yang salah. *Precision* negatif yang 81,33% menunjukkan bahwa model efektif dalam memprediksi kelas negatif. Dengan *recall* positif 75%, model dapat mendeteksi sebagian besar data positif. Sementara itu, *recall* negatif 84,72% menunjukkan bahwa model sangat baik dalam mendeteksi

data negatif. *F1-score* positif dan negatif yang masing-masing mencapai 77,06% dan 82,99% menunjukkan keseimbangan yang baik antara precision dan recall untuk kedua kelas. Akurasi sebesar 80,47% mengindikasikan bahwa model secara keseluruhan memprediksi dengan benar lebih dari 80% dari data yang diuji, meskipun masih ada ruang untuk perbaikan, terutama dalam mendeteksi kelas positif secara lebih optimal.

Tabel 4.10 Hasil 10-fold Cross Validation Rasio 80:20

n-fold	Precision Positif	Precision Negatif	Recall Positif	Recall Negatif	F1-score Positif	F1-score Negatif	Akurasi	Rata-Rata Akurasi
1	65,20%	75,90%	60%	79,80%	62,50%	77,80%	72,10%	75,20%
2	65,62%	84,62%	80,77%	71,43%	72,41%	77,46%	75,19%	
3	68,97%	78,73%	72,73%	75,68%	70,73%	77,19%	74,42%	
4	60,78%	82,05%	68,89%	76,19%	64,56%	79,04%	73,64%	
5	55,70%	83,80%	75,60%	67,90%	64,20%	74,90%	70,50%	
6	73,33%	81,16%	77,19%	77,78%	75,21%	79,43%	77,52%	
7	79,25%	81,33%	75%	84,72%	77,06%	82,99%	80,47%	
8	70,80%	77,80%	76,70%	72,10%	73,60%	74,80%	74,20%	
9	71,19%	82,61%	77,78%	77,03%	74,36%	79,72%	77,34%	
10	78,67%	74,70%	74,70%	78,67%	76,60%	76,60%	76,60%	

4.5.2 Hasil Pengujian Rasio 70:30

Pada pengujian dengan rasio data 70:30, model diuji menggunakan 10-fold cross-validation selama 50 epoch per fold. Hasil pada Tabel 4.16 hasil *training loss* dan *validation accuracy* dengan *fold* terbaik berada pada *fold* ke-7 dengan nilai *training loss* epoch pertama 1.4077 kemudian menurun seiring bertambahnya epoch hingga mencapai 0,2552 pada *epoch* ke-50. *Validation accuracy* pada *fold* ke-7 dimulai dengan tingkat akurasi 0.5982 pada *epoch* pertama dan terus meningkat dengan akurasi terakhir mencapai 0.8047 pada *epoch* ke-50.

Tabel 4.11 Hasil Train. Loss dan Val. Accuracy 70:30 dengan 10-Fold

n-fold	Epoch	Training Loss	Validation Accuracy
1	1	1.4184	0.6372
	2	1.1258	0.6637

<i>n-fold</i>	<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Accuracy</i>
	3	0.9471	0.6726
	4	0.8308	0.6991
	5	0.7548	0.6814
2	1	1.4163	0.5487
	2	1.1075	0.5752
	3	0.9426	0.6549
	4	0.8306	0.6460
	5	0.7663	0.6991
3	1	1.4187	0.6637
	2	1.0974	0.7080
	3	0.9030	0.7522
	4	0.7946	0.7434
	5	0.7277	0.7345
4	1	1.4224	0.5575
	2	1.1096	0.6460
	3	0.9218	0.6283
	4	0.8021	0.6991
	5	0.7249	0.7080
5	1	1.4131	0.4643
	2	1.0501	0.7411
	3	0.8552	0.7500
	4	0.7496	0.7500
	5	0.6816	0.7589
6	1	1.4273	0.4732
	2	1.1060	0.6429
	3	0.9374	0.6875
	4	0.8168	0.7857
	5	0.7464	0.7500
7	1	1.4077	0.5982
	2	1.0528	0.7946
	3	0.8600	0.8214
	4	0.7572	0.8304
	5	0.6975	0.7946
8	1	1.3976	0.6607
	2	1.0159	0.6429
	3	0.8179	0.6964
	4	0.7141	0.7768
	5	0.6535	0.7768
9	1	1.4157	0.5536
	2	1.0827	0.6339
	3	0.8820	0.7232
	4	0.7738	0.7768
	5	0.6921	0.7411
10	1	1.4049	0.6696
	2	1.0401	0.6607
	3	0.8466	0.6786
	4	0.7599	0.6429
	5	0.6852	0.7321

Hasil rasio pengujian 70:30 dengan menggunakan *10-fold cross validation* mendapatkan hasil perhitungan seperti pada Tabel 4.11. Evaluasi dilakukan dengan mengukur metrik klasifikasi seperti *precision*, *recall*, dan *F1-score* untuk masing-masing kelas positif dan negatif, serta nilai akurasi keseluruhan. Hasil menunjukkan bahwa nilai *precision* kelas positif berkisar antara 71,43% hingga 86,67%, sedangkan *precision* negatif antara 65,79% hingga 82,05%. Nilai *recall* untuk kelas positif mencapai titik tertinggi sebesar 88,89% pada *fold* ke-9, sementara *recall* negatif tertinggi sebesar 83,67% pada *fold* ke-7. *F1-score* terbaik untuk kedua kelas juga muncul pada *fold* ke-7, yakni 84,57% (positif) dan 81,23% (negatif), menunjukkan performa model paling seimbang pada *fold* tersebut.

Akurasi keseluruhan berada dalam rentang 72,68% hingga 83,04%, dengan rata-rata akurasi sebesar 78,34%. *Fold* terbaik pada rasio 70:30 berada pada *fold* ke 7 dengan nilai *precision* positif 86,67% , *precision* negatif 78,85%, *recall* positif 82,54%, *recall* negatif 83,67%, *F1-score* positif 84,57%, negatif 81,23% dan akurasi 83,04%. Hal ini berarti model menunjukkan performa yang sangat baik pada *fold* ini, dengan kemampuan yang sangat baik dalam memprediksi kelas positif dan negatif. *Precision* positif yang tinggi (86,67%) menunjukkan bahwa model sangat akurat dalam memprediksi kelas positif, dengan sedikit kesalahan (False Positives). *Precision* negatif sebesar 78,85% mengindikasikan bahwa model juga cukup baik dalam memprediksi kelas negatif, meskipun ada beberapa kesalahan pada prediksi negatif. *Recall* positif sebesar 82,54% menunjukkan bahwa model mampu mendeteksi sebagian besar data positif yang ada, meskipun ada sedikit data positif yang terlewat. Dengan *recall negatif* sebesar 83,67%, model

juga sangat efektif dalam mendeteksi data negatif. *F1-score* positif dan negatif yang masing-masing sebesar 84,57% dan 81,23% menunjukkan keseimbangan yang sangat baik antara precision dan recall untuk kedua kelas. Akurasi sebesar 83,04% menunjukkan bahwa model berhasil memprediksi dengan benar lebih dari 83%.

Tabel 4.12 Hasil 10-fold Cross Validation Rasio 70:30

n-fold	Precision Positif	Precision Negatif	Recall Positif	Recall Negatif	F1-score Positif	F1-score Negatif	Akurasi	Rata-Rata Akurasi
1	74,29%	69,77%	80%	62,50%	77,05%	65,93%	72,68%	78,34%
2	71,43%	74%	77,59%	67,27%	74,44%	70,53%	72,68%	
3	77,05%	73,08%	77,05%	73,08%	77,05%	73,08%	75,22%	
4	73,97%	72,50%	83,08%	60,42%	78,36%	65,92%	73,45%	
5	82,54%	67,35%	76,47%	75%	79,32%	71,03%	75,89%	
6	80,30%	73,91%	81,54%	72,34%	80,91%	73,12%	77,68%	
7	86,67%	78,85%	82,54%	83,67%	84,57%	81,23%	83,04%	
8	79,73%	65,79%	81,94%	62,50%	80,83%	64,19%	75%	
9	76,70%	82,05%	88,89%	65,31%	82,45%	72,82%	78,57%	
10	76,70%	66,67%	81,01%	60,47%	78,79%	63,49%	73,21%	

4.5.3 Hasil Evaluasi Model

Evaluasi performa model dilakukan dengan membandingkan dua rasio pembagian data dan pengujian menggunakan *k-fold cross-validation*, pada dua rasio pembagian data yang berbeda. Berdasarkan Tabel 4.9, akurasi model pada rasio 80:20 memiliki nilai rata – rata akurasi 75,20% dengan nilai akurasi tertinggi berada pada fold ke-7 mencapai nilai akurasi 80,47%, *precision* positif sebesar 79,25%, *precision* negatif 81,33%, *recall* positif 75%, *recall* negatif 84,72%, *F1-score* positif 77,06%, dan *F1-score* negatif 82,99%.

Model prediksi terbaik dalam penelitian ini berada pada pembagian data 70:30 dapat dilihat pada Tabel 4.12 perbandingan akurasi kedua rasio. Rasio 70:30 khususnya pada *fold* ke-7. Pada *fold* ini, model mencatat *precision positif* sebesar 86,67%, *precision negatif* 78,85%, *recall positif* 82,54%, *recall negatif* 83,67%,

F1-score positif 84,57%, F1-score negatif 81,23%, dan akurasi sebesar 83,04%. Nilai ini merupakan yang tertinggi dibandingkan *fold* lainnya, menunjukkan performa yang paling seimbang dan akurat dalam klasifikasi dua kelas. Hal ini menunjukkan bahwa penerapan *k-fold cross-validation* mampu meningkatkan stabilitas dan generalisasi model.

Tabel 4.13 Hasil Perbandingan Akurasi Kedua Skenario

Rasio	Akurasi rata-rata menggunakan K-fold
80:20	75,20%
70:30	78,34%

Setelah diperoleh model Bi-LSTM dengan performa terbaik berdasarkan hasil evaluasi sebelumnya, tahap selanjutnya adalah melakukan pengujian terhadap data uji untuk melihat ketepatan prediksi model. Pengujian ini bertujuan untuk membandingkan antara status pelunasan haji berdasarkan data aktual dan hasil prediksi model. Tabel 4.13 menyajikan hasil prediksi pelunasan haji menggunakan model terbaik Bi-LSTM yang telah dilatih sebelumnya dengan skenario pembagian data 70:30 dan penerapan *k-fold cross-validation*. Model ini merupakan hasil dari *fold* ke-7 yang menunjukkan performa terbaik dengan akurasi 83,04%, *precision* positif 86,67%, dan *F1-score* positif sebesar 84,57%.

Tabel 4.14 Perbandingan Prediksi Pelunasan Haji

Porsi	Pelunasan Haji Data Aktual	Prediksi Bi-LSTM
1300591006	Melunasi	Tidak Melunasi
1300591009	Melunasi	Melunasi
1300591016	Melunasi	Tidak Melunasi
1300591034	Tidak Melunasi	Tidak Melunasi
1300590631	Melunasi	Melunasi
1300561619	Melunasi	Melunasi
1300561969	Tidak Melunasi	Tidak Melunasi
1300562036	Melunasi	Melunasi
1300562310	Melunasi	Melunasi
1300562311	Melunasi	Melunasi

Berdasarkan data aktual, terdapat 319 jamaah calon haji yang benar-benar melunasi dan 158 jamaah yang benar-benar tidak melunasi. Tabel 4.15 memperlihatkan confusion matrix dari hasil pengujian model Bi-LSTM pada data uji. Berdasarkan visualisasi tersebut, sebanyak 235 data termasuk dalam kategori *true positive* (melunasi dan diprediksi melunasi), serta 136 data termasuk dalam kategori *true negative* (tidak melunasi dan diprediksi tidak melunasi). Sementara itu, terdapat 84 data yang tergolong *false negative* (melunasi namun diprediksi tidak melunasi) dan 22 data yang tergolong *false positive* (tidak melunasi namun diprediksi melunasi). Secara keseluruhan model memprediksi sebanyak 66,8% jamaah calon haji diprediksi dapat melunasi, dan sebanyak 33,12 % jamaah calon haji diprediksi tidak dapat melunasi. Dari jumlah tersebut, model Bi-LSTM berhasil memprediksi sesuai dengan data aktual sebanyak 235 jamaah yang melunasi (73,67% dari 319) dan 136 jamaah yang tidak melunasi (86,08% dari 158).

Tabel 4.15 Hasil *confusion matrix* model terbaik

Prediksi	Data Aktual	
	Positif	Negatif
Positif	235	22
Negatif	84	136

4.6 Pembahasan

Dari hasil penelitian ini menggunakan metode Bi-LSTM untuk prediksi pelunasan biaya haji di Kemenag Kota Malang, Dengan jumlah data sebanyak 1.619 data pelunasan haji, yang diberi label melunasi dan tidak melunasi, terdapat 920 data melunasi dan 699 data tidak melunasi. Kemudian dilakukan *preprocessing* untuk pembersihan data. Kemudian membuat model pelatihan menggunakan split

data dengan menggunakan *k-fold cross validation*. Terakhir yaitu mengevaluasi dengan menggunakan *confusion matrix* dan menghitung matrix seperti akurasi, *precision*, *recall* dan *f1-score*.

Split data yang digunakan yaitu 70:30 dan 80:20. Dari hasil pelatihan model split data menggunakan *k-fold* dengan $k=10$. Mendapatkan hasil dengan akurasi terbaik juga berada pada rasio 70:30 dengan nilai akurasi sebesar 83,04%. Hal ini menunjukkan bahwa metode split data menggunakan *k-fold* lebih signifikan untuk melakukan analisis prediksi.

Tujuan dari penelitian ini adalah menerapkan model Bi- LSTM untuk memprediksi pelunasan haji berdasarkan data historis pada Kantor Kementerian Agama Kota Malang. Meskipun penelitian ini belum secara langsung diimplementasikan oleh pihak Kementerian Agama Kota Malang, namun secara akademis, penelitian ini diharapkan dapat memberikan kontribusi ilmiah terhadap pengembangan sistem prediksi dalam bidang pelayanan haji, khususnya dari sisi analisis data dan teknologi kecerdasan buatan.

Dalam Islam, bekerja dengan kesungguhan dan menyempurnakan tanggung jawab adalah bagian dari nilai ihsan dalam beramal. Hal ini ditegaskan dalam sabda Rasulullah SAW:

إِنَّ اللَّهَ عَزَّ وَجَلَّ يُحِبُّ إِذَا عَمِلَ أَحَدُكُمْ عَمَلًا أَنْ يُتْقِنَهُ

Dari Aisyah r.a., sesungguhnya Rasulullah s.a.w. bersabda: “Sesungguhnya Allah mencintai seseorang yang apabila bekerja, mengerjakannya secara itqan (professional)”. (HR. Thabrani).

Hadis ini menekankan bahwa setiap bentuk pekerjaan hendaknya dilakukan dengan ketelitian, kesungguhan, dan tanggung jawab yang tinggi. Penggunaan

model Bi- LSTM dalam penelitian ini dapat dipahami sebagai bentuk implementasi nilai itqan. Namun demikian, Islam juga memberikan batasan yang jelas mengenai kemampuan manusia dalam mengetahui hal-hal yang berkaitan dengan masa depan. Firman Allah SWT dalam tafsir ibnu Katsir QS. Yusuf ayat 76 (Al-Sheikh, 2010):

فَبَدَأَ بِأَوْعِيَّتِهِمْ قَبْلَ وِعَاءِ أَخِيهِ ثُمَّ اسْتَخْرَجَهَا مِنْ وِعَاءِ أَخِيهِ كَذَلِكَ كِدْنَا لِيُوسُفَ مَا كَانَ لِيَأْخُذَ أَخَاهُ فِي دِينِ الْمَلِكِ إِلَّا أَنْ يَشَاءَ اللَّهُ نَرْفَعُ دَرَجَاتٍ مَنْ نَشَاءُ وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ ﴿٧٦﴾

“Maka, mulailah dia (memeriksa) karung-karung mereka sebelum (memeriksa) karung saudaranya sendiri (Bunjamin), kemudian dia mengeluarkannya (cawan raja itu) dari karung saudaranya. Demikianlah Kami mengatur (rencana) untuk Yusuf. Dia tidak dapat menghukum saudaranya menurut hukum raja, kecuali Allah menghendakinya. Kami angkat derajat orang yang Kami kehendaki; dan di atas setiap orang yang berpengetahuan ada yang lebih mengetahui.” (QS. Yusuf ayat 76)

Ayat Al-Quran surat Yusuf ayat 76 memberikan batas yang jelas antara pengetahuan manusia dan pengetahuan Allah. Pada Ayat ini memberikan penegasan bahwa hasil prediksi yang bersifat statistik atau berbasis teknologi tidak bisa diposisikan sebagai sesuatu yang absolut, melainkan sebagai bentuk ikhtiar manusia berdasarkan data yang ada. Meskipun manusia diperbolehkan melakukan estimasi dan prediksi berdasarkan data dan logika, hasil akhirnya tetap berada dalam kuasa Allah. Oleh karena itu, prediksi dalam penelitian ini diposisikan sebagai alat bantu keputusan, bukan sebagai penentu mutlak. Ini menegaskan bahwa ikhtiar ilmiah tidak bertentangan dengan tauhid, selama tetap dalam batas etika dan kesadaran akan keterbatasan manusia.

Dengan demikian, penelitian ini tidak hanya menghasilkan kontribusi pada aspek teknologi informasi, tetapi juga memiliki nilai spiritual yang tinggi karena

menjadi bagian dari upaya membantu terselenggaranya ibadah dengan lebih baik, adil, dan tepat sasaran.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk menerapkan metode Bi-LSTM dalam memprediksi pelunasan biaya haji berdasarkan data historis calon jemaah dari Kementerian Agama Kota Malang. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa model Bi-LSTM mampu melakukan klasifikasi status pelunasan dengan performa model baik (Geografi et al., 2022). Hal ini dibuktikan dengan akurasi tertinggi yang dicapai sebesar 83,04% pada skenario pembagian data 70:30 pada fold ke-7 menggunakan *10-fold cross-validation*.

Model yang dibangun berhasil memanfaatkan variabel demografis dan temporal untuk mengidentifikasi pola pelunasan secara efektif. Dengan demikian, penerapan metode Bi-LSTM dalam penelitian ini menunjukkan potensi yang kuat dalam mendukung sistem prediksi pelayanan haji, khususnya dalam membantu pengambilan keputusan secara lebih tepat, efisien, dan berbasis data.

5.2 Saran

Untuk pengembangan lebih lanjut pada penelitian yang telah dilakukan pada sistem prediksi menggunakan metode Bi-LSTM dalam memprediksi pelunasan biaya haji:

1. Penelitian ini masih menggunakan arsitektur Bi-LSTM yang sederhana dengan satu *hidden layer*. Untuk penelitian selanjutnya, disarankan agar model dikembangkan dengan menambahkan jumlah *hidden layer* agar

mampu mempelajari pola data yang lebih dalam dan kompleks, terutama jika jumlah data pelatihan ditingkatkan atau variasi fitur semakin banyak. Penambahan *hidden layer* dapat membantu model menangkap representasi yang lebih abstrak dari hubungan antar fitur prediktor.

2. Penelitian ini sudah menggunakan metrik *precision*, *recall*, dan *F1-score*. Namun, untuk memperkaya evaluasi performa model juga dapat diperluas dengan menambahkan AUC-ROC (*Area Under the Receiver Operating Characteristic Curve*) dan *Log Loss*. AUC-ROC memberikan gambaran kemampuan model dalam membedakan antara kelas melunasi dan tidak melunasi secara keseluruhan, sementara *Log Loss* mengukur seberapa baik probabilitas prediksi model terhadap label aktual. Dengan demikian, kelemahan prediksi yang tidak tertangkap oleh metrik klasifikasi konvensional dapat teridentifikasi secara lebih menyeluruh dan numerik.

DAFTAR PUSTAKA

- Ajayaghosh, N. (2024). AI Assisted Deceptive Content Analysis System Using Bi-LSTM. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2024.62143>
- Al-Sheikh, A. bin M. bin A. bin I. (2010). *Tafsir Ibnu Katsir Surat Yusuf* (pp. 1–74).
- Anas, K., Nurhayati, W., & Aslichah. (2023). Evaluasi Kebijakan Penyelenggaraan Haji dan Dampaknya Terhadap Pengeluaran Pemerintah Indonesia Tahun 2023. *Mutiara: Jurnal Ilmiah Multidisiplin Indonesia*, 1(1), 170–183. <https://doi.org/10.61404/jimi.v1i1.50>
- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(April), 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Binbusayyis, A., & Sha, M. (2025). Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism. *Heliyon*, 11(1). <https://doi.org/10.1016/j.heliyon.2024.e41507>
- Dewi, M. N., & Putra, R. E. (2024). Pengembangan Sistem Analisis Sentimen Masyarakat Terhadap Chatgpt Pada Twitter Dengan Perbandingan Metode Naive Bayes Classifier Dan K-Nearest Neighbors. *Journal of Informatics and Computer Science (JINACS)*, 5(03), 344–357. <https://doi.org/10.26740/jinacs.v5n03.p344-357>
- El-Kafrawy, S. A., Alsayed, S. M., Faizo, A. A. A., Bajrai, L. H., Uthman, N. A., Alsaheed, M., Hassan, A., Alquthami, K. M., Alandijany, T. A., Zumla, A., & Azhar, E. I. (2023). Genetic diversity and molecular analysis of human influenza virus among pilgrims during Hajj. *Heliyon*, 10. <https://doi.org/10.1016/j.heliyon.2023.e23027>
- Endang Etriyanti. (2021). Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 3(1), 6–14. <https://doi.org/10.52303/jb.v3i1.40>
- Geografi, J., Ilmu, F., Unp, S., Lima, K., & Kota, P. (2022). *ANALISIS KERAWANAN LONGSOR MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON DI KECAMATAN SULIKI KABUPATEN LIMA PULUH KOTA*. 4.
- Hu, X., Yuan, S., Xu, F., Leng, Y., Yuan, K., & Yuan, Q. (2020). Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Computers in Biology and Medicine*, 124(July), 103919. <https://doi.org/10.1016/j.compbiomed.2020.103919>
- Islami Rahayu, N. W., Fatimatuazzahro, F., Rakhman Wijaya, A., Mutmainnah, M.,

- Putra Nusantara, A. F., & Ahmad Kurniawan, Z. (2024). Memprediksi Persepsi Umat Muslim Indonesia Terhadap Kebijakan Pemerintah Terkait Kenaikan Biaya Penyelenggaraan Haji. *Majalah Ilmiah Dian Ilmu*, 24(1), 47–66. <https://doi.org/10.37849/midi.v24i1.403>
- Justin, S., Wao, A., & Wao, A. A. (2024). Performance Analysis of Activation Functions in Neural Networks. *ShodhKosh: Journal of Visual and Performing Arts*, 5(1), 588–598. <https://doi.org/10.29121/shodhkosh.v5.i1.2024.190>
- K, N., M, K., Easwaramoorthy, S. V., C R, D., Yoo, S., & Cho, J. (2024). Hybrid approach of deep feature extraction using BERT– OPCNN & FIAC with customized Bi-LSTM for rumor text classification. *Alexandria Engineering Journal*, 90(January), 65–75. <https://doi.org/10.1016/j.aej.2024.01.056>
- Kusuma, A. B., & Hadinata, N. (2022). The Implementation of the Black Box Method for Testing Smart Hajj Application Ministry of Religion. *Journal of Information Systems and Informatics*, 4(3), 673–686. <https://doi.org/10.51519/journalisi.v4i3.306>
- Lasijan, T. G., Santoso, R., & Hakim, A. R. (2023). Prediksi Harga Emas Dunia Menggunakan Metode Long-Short Term Memory. *Jurnal Gaussian*, 12(2), 287–295. <https://doi.org/10.14710/j.gauss.12.2.287-295>
- N., P., R., K., A., K., & Kakani, V. (2023). A Bidirectional LSTM approach for written script auto evaluation using keywords-based pattern matching. *Natural Language Processing Journal*, 5(May), 100033. <https://doi.org/10.1016/j.nlp.2023.100033>
- Nyoman, S. G. (2020). Pengaruh Tingkat Pendidikan dan Jenis Pekerjaan Terhadap Produktivitas dan Pendapatan Keluarga Miskin Kecamatan Negara Jembrana. *E-Jurnal EP Unud*, 9, 2529–2556.
- Philip Thekkekara, J., Yongchareon, S., & Liesaputra, V. (2024). An attention-based CNN-BiLSTM model for depression detection on social media text. *Expert Systems with Applications*, 249(PC), 123834. <https://doi.org/10.1016/j.eswa.2024.123834>
- Phillips, G., Teixeira, H., Kelly, M. G., Salas Herrero, F., Várbró, G., Lyche Solheim, A., Kolada, A., Free, G., & Poikane, S. (2024). Setting nutrient boundaries to protect aquatic communities: The importance of comparing observed and predicted classifications using measures derived from a confusion matrix. *Science of The Total Environment*, 912, 168872. <https://doi.org/10.1016/j.scitotenv.2023.168872>
- Qolbi, L., & Rahman, A. (2023). The Effectiveness of SISKOHAT in Organizing Hajj at the Ministry of Religious Affairs Office of Lima Puluh Kota Regency. *Munazzama: Journal of Islamic Management and Pilgrimage*, 3(2), 140–150. <https://doi.org/10.21580/mz.v3i2.17226>
- Sofianiyatin, L., & Pradana, A. R. (2024). PENERAPAN TABUNGAN IB

HIJRAH RENCANA SEBAGAI ALTERNATIF PELUNASAN BIAYA
HAJI PADA PT. BANK MUAMALAT INDONESIA, KANTOR CABANG
PEMBANTU MADIUN. *Commodity*, 3(1), 33–42.
<https://doi.org/10.56997/commodity.v3i1.1413>

Tollner, D., Ziyu, W., Zöldy, M., & Török, Á. (2024). Demonstrating a new evaluation method on ReLU based Neural Networks for classification problems. *Expert Systems with Applications*, 250(February).
<https://doi.org/10.1016/j.eswa.2024.123905>

Tran, D. T., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2021). Data Normalization for Bilinear Structures in High-Frequency Financial Time-series. *2020 25th International Conference on Pattern Recognition (ICPR)*, 7287–7292. <https://doi.org/10.1109/ICPR48806.2021.9412547>

Wardatur Rizqiyah, S. A., & Chonyta, D. (2024). Efektivitas Sistem Informasi dan Komputerisasi Haji Terpadu (SISKOHAT) dalam Pelayanan Pendaftaran Calon Jemaah Haji Reguler (Studi di Kantor Pusat Layanan Haji dan Umrah Terpadu Kraksaan). *Innovative: Journal Of Social Science Research*, 4(3), 10886–10896. <https://doi.org/10.31004/innovative.v4i3.11685>

Zen, A. F. A., Pramukantoro, E. S., Amron, K., Wardhani, V., & Kamila, P. A. (2024). Prediksi Detak Jantung Berbasis LSTM pada Raspberry Pi untuk Pemantauan Kesehatan Portabel. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(5), 1097–1104. <https://doi.org/10.25126/jtiik.2024118015>

LAMPIRAN

LAMPIRAN – Permohonan Penelitian



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jalan Gajayana 50 Malang 65144 Telepon/Faksimile (0341) 558933
Website: <http://saintek.uin-malang.ac.id>, email: saintek@uin-malang.ac.id

Nomor : B-97.O/FST.01/TL.00/08/2024
Lampiran : -
Hal : Permohonan Penelitian

Yth. Pimpinan Kantor Kementerian Agama Kota Malang
Jl. Raden Panji Suroso No.2, Polowijen, Kec. Blimbing, Kota Malang, Jawa Timur 65126

Dengan hormat,

Sehubungan dengan penelitian mahasiswa Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang atas nama:

Nama : NURJIHAN NABILAH RAMADHANI
NIM : 210605110061
: Implementasi Metode Bi-LSTM untuk Menganalisis Pengaruh
Judul Penelitian Kebijakan Pemerintah terhadap Pelunasan Biaya Haji di Kemenag Kota
Malang
Dosen : SUPRIYONO,M.Kom
Pembimbing

Maka kami mohon Bapak/Ibu berkenan memberikan izin pada mahasiswa tersebut untuk melakukan penelitian di Kantor Kementerian Agama Kota Malang dengan waktu pelaksanaan pada tanggal 12 Agustus 2024 sampai dengan 12 September 2024.

Demikian permohonan ini, atas perhatian dan kerjasamanya disampaikan terimakasih.

Malang, 23 Agustus 2024
a.n Dekan

Scan QRCode ini



untuk verifikasi surat



Wakil Dekan Bidang Akademik,

Dr Anton Prasetyo, M.Si
NIP. 19770925 200604 1 003