TEKNIMEDIA - Volume 7, Nomor 1, Juni 2026: 250 - 262

PREDIKSI TARGET PENDAPATAN PAJAK DAERAH DI KABUPATEN SUMBAWA MENGGUNAKAN ALGORITMA *EXTREME GRADIENT BOOSTING* (XGBOOST)

(*PREDICTION OF LOCAL TAX REVENUE TARGET IN SUMBAWA REGENCY USING THE EXTREME GRADIENT BOOSTING ALGORITHM* (XGBOOST))

Sukarti[1)], Ekastini[2)]
[1,2)]Informatika Universitas Teknologi Sumbawa
Jl. Raya Olat Maras Batu Alang, Pernek, Kec. Moyo Hulu, Kabupaten Sumbawa, Nusa Tenggara Barat
e-mail: sukartisumbawa692@gmail.com[1)], ekastini@uts.ac.id[2)]

ABSTRAK

Pendapatan Asli Daerah (PAD) merupakan sumber pembiayaan utama pemerintah daerah, di mana pajak daerah menjadi komponen terbesar dalam penyokong penerimaan. Di Kabupaten Sumbawa, penetapan target pajak masih bertumpu pada potensi dan realisasi historis sehingga ketepatan dalam menetapkan target pajak belum optimal. Penelitian ini bertujuan untuk membangun model prediksi target pendapatan pajak daerah menggunakan algoritma XGBoost. Metode penelitian yang digunakan adalah kerangka kerja Cross Industry Standard Process For Data mining (CRISP-DM) dengan tahapan business understanding, data understanding, preprocessing, modelling, evaluation, dan deployment. Data yang digunakan dalam penelitian ini merupakan data sekunder diperoleh dari Bapenda Kabupaten Sumbawa, mencakup berbagai jenis pajak daerah pada periode 2021–2025. Hasil evaluasi menunjukkan nilai RMSE sebesar 777,824,418.79 atau 777 juta, MAPE 5,14%, dan R^2 98% yang menunjukkan kesalahan prediksi rendah serta kemampuan model memahami pola data dengan baik. Model kemudian diimplementasikan ke dalam sistem web berbasis flask untuk mendukung proses input data data, proses pelatihan model (training) serta proses memprediksi target pendapatan pajak pajak daerah secara lebih akurat dan berbasis data. Sistem ini diharapkan dapat menjadi alat bantu strategis bagi pemerintah daerah dalam pengambilan keputusan terkait penetapan target pajak daerah.

Kata Kunci: *Pajak Daerah, XGBoost, Prediksi, CRISP-DM, Machine Learning*

ABSTRACT

Locally Generated Revenue (PAD) is the primary source of financing for local governments, with local taxes being the largest component in revenue generation. In Sumbawa Regency, tax target determination is still based on historical potential and realization, resulting in suboptimal accuracy in tax target determination. This study aims to develop a prediction model for local tax revenue targets using the XGBoost algorithm. The research method used is the Cross Industry Standard Process for Data Mining (CRISP-DM) framework with stages of business understanding, data understanding, preprocessing, modeling, evaluation, and implementation. The data used in this study is secondary data obtained from the Sumbawa Regency Regional Revenue Agency (Bapenda), covering various types of local taxes for the 2021–2025 period. The evaluation results show an RMSE value of 777,824,418.79 or 777 million, MAPE 5.14%, and R^2 98%, indicating low prediction errors and the model's ability to understand data patterns well. The model was then implemented into a Flask-based web system to support data input, model training, and more accurate, data-driven prediction of regional tax revenue targets. This system is expected to serve as a strategic tool for regional governments in decision-making regarding setting regional tax targets..

Keywords : *Regional Tax, XGBoost, Prediction, CRISP-DM, Machine Learning*

I. PENDAHULUAN

Pendapatan Asli Daerah (PAD) adalah penerimaan yang bersumber dari potensi ekonomi yang dimiliki oleh daerah itu sendiri [1]. Menurut Undang-Undang nomor 33 tahun 2004 tentang Perimbangan Keuangan Antara Pusat Dan Daerah pasal 1 angka 18 menyatakan bahwa PAD adalah pendapatan yang

diperoleh daerah yang dipungut berdasarkan peraturan daerah sesuai dengan peraturan perudang undangan [2]. Salah satu sumber penerimaan PAD adalah pajak daerah [3]. Berdasarkan Undang-Undang Nomor 28 Tahun 2009 tentang Pajak Daerah dan Retribusi Daerah (PDRD), Pajak daerah adalah kewajiban yang harus dibayarkan oleh individu maupun badan usaha kepada pemerintah daerah, bersifat memaksa tanpa adanya imbalan secara langsung, dan digunakan untuk kepentingan masyarakat secara luas [4].

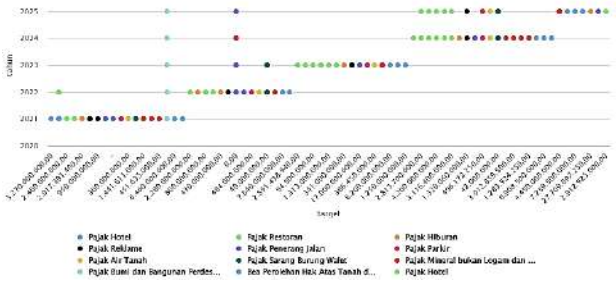
Dalam pengelolaan keuangan daerah, penentuan target pendapatan pajak daerah menjadi langkah penting karena ketepatan target akan memengaruhi efektivitas penyusunan Anggaran Pendapatan dan Belanja Daerah (APBD)[3]. Kerena Target yang realistis dan berbasis data dapat membantu optimalisasi penerimaan serta mendukung pembangunan yang berkelanjutan. Di Kabupaten Sumbawa, proses penetapan target pendapatan pajak daerah menggunakan dua pendekatan, yaitu metode potensi dan metode realisasi. Metode potensi digunakan untuk memperkirakan kemampuan sektor pajak, sedangkan metode realisasi mengacu pada perbandingann capaian pajak lima tahun terakhir diperoleh rata-rata serta persentase capaian yang menjadi dasar penetapan target pajak untuk tahun berikutnya.

Namun, kedua metode tersebut belum sepenuhnya optimal karena keterbatasan data historis yang menyebabkan hasil perhitungan target belum mencerminkan potensi riil daerah. Padahal, penetapan target pajak yang akurat sangat penting karena menjadi dasar penyusunan APBD dan perhitungan PAD yang digunakan untuk pembiayaan pembangunan.

Tabel 1. Target dan realisasi pajak daerah 2021-2025

Tahun	Target Pajak Daerah	Realisasi	%
2021	37.813.000.000,00	36.331.561.589,00	96,08%
2022	44.795.000.000,00	42.878.103.476,25	95,72%
2023	78.074.934.520,00	47.353.212.371,00	60,65%
2024	48.239.429.750,00	55.505.811.651,50	115,06
2025	81.744.248.345,00	59.958.731.084,88	73,35%

Selain tabel target dan realisasi pajak daerah berikut grafik target pendapatan pajak daerah 2021-2025.



Gambar 1. Grafik Target Pendapatan Pajak Daerah 2021-2025

Berdasarkan grafik target pendapatan pajak Daerah Kabupaten Sumbawa tahun 2021–2025, terlihat bahwa target pendapatan pajak daerah cenderung mengalami peningkatan setiap tahun pada sebagian besar jenis pajak. Beberapa jenis pajak seperti Pajak Restoran, Pajak Penerangan Jalan, dan BPHTB menunjukkan nilai target yang lebih tinggi dibandingkan jenis pajak lainnya, yang menandakan kontribusi besar terhadap Pendapatan Asli Daerah (PAD). Namun, terdapat perbedaan nilai target antar jenis pajak yang menunjukkan bahwa potensi penerimaan pajak daerah tidak merata.

Kondisi tersebut menunjukkan bahwa proses penetapan target pajak masih belum sepenuhnya sejalan dengan potensi ekonomi daerah. Sehingga dibutuhkan pendekatan yang lebih inovatif dan berbasis teknologi untuk meningkatkan ketepatan dalam menetapkan target pajak. Salah satu metode yang dapat diterapkan adalah dengan memanfaatkan metode *machine learning*, khususnya algoritma *Extreme Gradient Boosting* (XGBoost). XGBoost merupakan algoritma berbasis *gradient boosting* yang memiliki kemampuan tinggi dalam mengolah data kompleks, termasuk data deret waktu (*time series*), mempelajari pola non-linear dari data pajak secara efisien dan akurat [4].

Penelitian oleh [5] yang berjudul Implementasi Algoritma *XGBoost* Untuk

Prediksi Capaian Bulanan Pendapatan Daerah Kota Bandung menunjukkan bahwa algoritma XGBoost mampu menghasilkan tingkat akurasi yang tinggi dengan nilai *Mean Absolute Error* (MAE) sebesar Rp5,6 miliar, *Root Mean Square Error* (RMSE) sebesar Rp9,3 miliar, dan *koefisien determinasi* (R^2) sebesar 73% pada data uji. Selain itu, hasil validasi silang (*cross-validation*) 5-fold menunjukkan nilai R^2 sebesar 86%, yang menegaskan stabilitas model dalam melakukan prediksi.

Adapun penelitian sebelumnya yang di lakukan oleh [6] bertujuan untuk menganalisis data pajak dan retribusi daerah guna mengetahui pola pengelompokan data penerimaan pajak daerah. Metode yang digunakan dalam penelitian tersebut adalah algoritma K-Means, yaitu metode *clustering* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tertentu. Hasil penelitian menunjukkan bahwa data pajak dan retribusi daerah dapat dikelompokkan ke dalam beberapa kategori berdasarkan tingkat penerimaan sehingga memberikan gambaran mengenai distribusi dan pola data pajak daerah. Namun, kekurangan dari penelitian tersebut adalah pendekatan yang digunakan masih bersifat analisis deskriptif karena hanya berfokus pada proses pengelompokan data dan belum mampu melakukan prediksi terhadap nilai pajak di masa mendatang. Oleh karena itu, penelitian berjudul "Prediksi Target Pendapatan Pajak Daerah di Kabupaten Sumbawa Menggunakan Algoritma *Extreme Gradient Boosting* (XGBoost)" dilakukan untuk mengatasi keterbatasan tersebut dengan menerapkan metode *machine learning* yang berfokus pada prediksi. Penelitian ini diharapkan mampu memberikan model prediksi yang lebih akurat sehingga dapat membantu pemerintah daerah dalam menetapkan target pendapatan pajak secara lebih efektif, berbasis data, serta mendukung pengelolaan keuangan daerah yang transparan, akuntabel, dan berkelanjutan.

II. STUDI PUSTAKA

Beberapa penelitian terdahulu telah mengembangkan berbagai metode untuk memprediksi target pendapatan pajak daerah. Penelitian sebelumnya yang dilakukan oleh [7] yang berjudul Prediksi Jumlah Target dan Realisasi Wajib Pajak Atas PBB – P2 Menggunakan *Multi Regression*, *Regression Lasso*, dan *PCR.* Penelitian tersebut bertujuan untuk memprediksi jumlah target dan realisasi wajib pajak PBB-P2 dengan membandingkan tiga metode regresi, yaitu *Multi Regression*, *Regression Lasso*, dan *Principal Component Regression* (PCR). Hasil penelitian menunjukkan bahwa metode *Regression Lasso* dengan proses *pre-processing* memiliki kinerja terbaik dengan tingkat akurasi prediksi sebesar 79,08%, lebih unggul dibandingkan metode lainnya.

Penelitian lain yang dilakukan oleh [8] yang berjudul *Forecasting Sidoarjo Local Tax Revenue Using Exponential Smoothing Models* menggunakan metode *time series forecasting* untuk memprediksi Pendapatan Asli Daerah (PAD). Hasil penelitian menunjukkan bahwa metode peramalan mampu mengidentifikasi tren penerimaan pajak daerah sehingga dapat membantu pemerintah daerah dalam menyusun perencanaan anggaran secara lebih terukur.

Selanjutnya, penelitian yang dilakukan oleh [9] yang berjudul *Forecasting Value-Added Tax (VAT) Revenue Using Autoregressive Integrated Moving Average (ARIMA) Box-Jenkins Method.* Penelitian ini bertujuan membangun model peramalan penerimaan Pajak Pertambahan Nilai (Value-Added Tax/VAT) di Indonesia menggunakan metode ARIMA Box–Jenkins berdasarkan data deret waktu penerimaan pajak selama lima tahun terakhir dari dua Kantor Pelayanan Pajak Direktorat Jenderal Pajak. Hasil penelitian menunjukkan bahwa model ARIMA mampu menghasilkan prediksi penerimaan VAT yang lebih mendekati nilai realisasi dibandingkan dengan target penerimaan pajak yang ditetapkan pemerintah. Model peramalan ini dapat digunakan sebagai alat pendukung dalam menentukan target penerimaan pajak yang lebih realistis serta

sebagai indikator evaluasi kinerja Direktorat Jenderal Pajak.

Penelitian berikutnya dilakukan oleh [10] yang berjudul Optimalisasi Metode *Conjugate Gradient Polak Rebiere* Untuk Memprediksi Target PBB Kecamatan Dolok Merawan. Penelitian ini mengoptimalkan metode *Conjugate Gradient Polak Rebiere* untuk memprediksi target PBB di Kecamatan Dolok Merawan menggunakan algoritma Jaringan Syaraf Tiruan (JST). Penelitian ini menghasilkan *Mean Square Error* (MSE) sebesar 0,001649442 dengan akurasi prediksi mencapai 88% menggunakan arsitektur 5-4-1. Hasil penelitian membuktikan bahwa metode JST *Conjugate Gradient Polak Rebiere* dapat digunakan baik melalui perhitungan manual maupun menggunakan software Matlab 2011 untuk memprediksi target realisasi PBB.

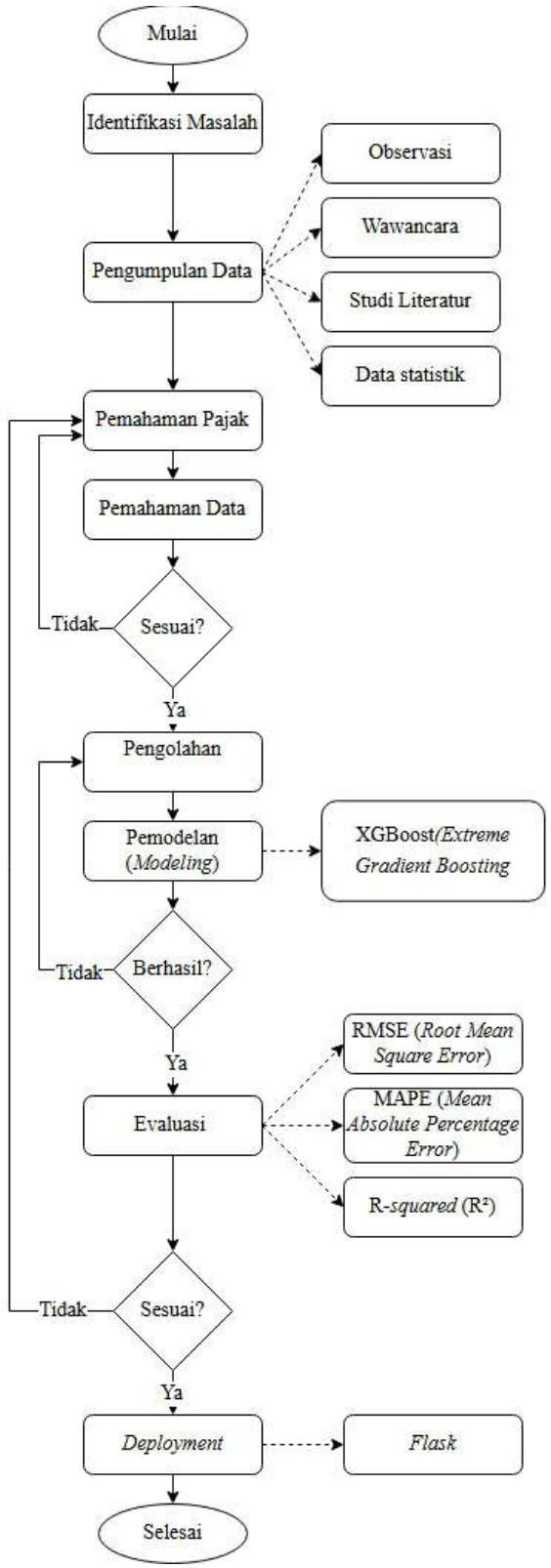
Selain itu, penelitian yang dilakukan oleh [11] yang berjudul Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. Penelitian ini membandingkan kinerja tiga algoritma, yaitu Support Vector Machine (SVM), Random Forest, dan XGBoost. Hasil penelitian dengan menggunakan algoritma SVM, random forest, dan XGBoost mendapatkan nilai accuracy, recall, precision tertinggi pada model XGBoost dengan nilai accuracy sebesar 82%, recall 70%, dan precision 92%. Hal ini menunjukkan bahwa algoritma XGBoost memiliki kinerja yang lebih baik dibandingkan algoritma SVM dan Random Forest.

Berdasarkan tinjauan pustaka di atas, dapat disimpulkan bahwa berbagai metode prediksi telah diterapkan untuk meramalkan penerimaan pajak daerah dengan tingkat keberhasilan yang bervariasi. Metode machine learning seperti *Regression Lasso*, XGBoost, dan Jaringan Syaraf Tiruan menunjukkan performa yang baik dengan tingkat akurasi yang tinggi, berkisar antara 79% hingga 88%. Sementara itu, metode statistik tradisional seperti *Single Moving Average* dan Model *Grey-Markov* juga masih relevan digunakan dalam konteks tertentu, meskipun dengan tingkat akurasi yang bervariasi.

Berdasarkan hasil kajian tersebut, penelitian ini akan mengembangkan model prediksi target pendapatan pajak daerah di Kabupaten Sumbawa dengan menerapkan algoritma XGBoost, yang dikenal efektif dalam memproses data deret waktu (*time series*). Hasil dari model prediksi ini akan diimplementasikan ke dalam bentuk *web* interaktif berbasis *flask*, yang menyajikan proses tambahkan data, train model, serta hasil prediksi untuk membantu pemerintah daerah dalam pengambilan keputusan yang lebih tepat terkait penetapan target pajak.

III. METODE PENELITIAN

Dalam penelitian ini, peneliti menerapkan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai kerangka kerja dalam proses pengolahan dan analisis data dan algoritma yang digunakan dalam penelitian ini adalah *Extreme Gradient Boosting* (XGBoost) untuk menghasilkan model prediksi yang lebih akurat berdasarkan pola data historis. Untuk mempermudah pelaksanaan penelitian, peneliti menyusun alur penelitian sebagai gambaran tahapan yang akan dilakukan selama proses penelitian. Alur penelitian dapat dilihat pada gambar 2 berikut ini.



Gambar 2. Alur penelitian

A. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan mengintegrasikan beberapa teknik pengumpulan data untuk memperoleh informasi secara menyeluruh. Metode yang digunakan meliputi observasi, wawancara, studi literatur, serta pemanfaatan dataset statistik terkait target pendapatan pajak daerah.

1. Studi Literatur

   Studi literatur digunakan untuk memperkuat pemahaman teoritis terkait konsep pendapatan pajak daerah, metode prediksi, dan algoritma XGBoost. Teknik ini juga mendukung penyusunan hipotesis serta perumusan metodologi penelitian yang sejalan dengan tujuan penelitian.

2. Observasi

   Observasi dilakukan secara langsung pada instansi terkait, khususnya di Badan Pendapatan Daerah (Bapenda) Kabupaten Sumbawa, guna memperoleh informasi empiris mengenai proses pengelolaan target pendapatn pajak daerah. Melalui observasi ini, peneliti dapat memahami kondisi aktual, mekanisme pencatatan, serta faktor-faktor yang memengaruhi capaian target pendapatan pajak daerah di wilayah tersebut.

3. Wawancara

   Wawancara dilakukan dengan Bapak Syamsul Bahri, S.E selaku kepala bidang tiga (perencanaan pengembangan dan pelaporan pendapatan daerah) di Bapenda Kabupaten Sumbawa. Wawancara bertujuan menggali informasi mendalam mengenai mekanisme penetapan target pajak, faktor pendukung dan penghambat capaian target, serta strategi operasional dalam pengelolaan pendapatan pajak daerah. Informasi yang diperoleh digunakan sebagai landasan dalam memahami konteks permasalahan dan pengembangan model prediksi yang relevan.

4. *Dataset* Statistik

   Dataset statistik diperoleh dari Bapenda Kabupaten Sumbawa yang berisi data historis target dan realisasi pendapatan pajak daerah beberapa tahun

terakhir. Dataset ini menjadi sumber utama dalam analisis kuantitatif, pemodelan, dan evaluasi algoritma XGBoost.

5. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang berasal dari dokumentasi pemerintah terkait statistik pendapatan pajak daerah Kabupaten Sumbawa tahun 2021-2025. Jenis pajak yang di gunakan dalam penelitian ini meliputi Pajak Hotel, Pajak Restoran, Pajak Hiburan, Pajak Reklame, Pajak Penerangan Jalan, Pajak Mineral Bukan Logam dan Batuan, Pajak Parkir, Pajak Air Tanah, Pajak Sarang Burung Walet, Pajak Bumi dan Bangunan Perdesaan dan Perkotaan (PBB-P2), serta Bea Perolehan Hak atas Tanah dan Bangunan (BPHTB).

Dataset yang digunakan dalam penelitian ini berjumlah 376 data yang merupakan data historis penerimaan pajak daerah dan digunakan sebagai dasar dalam membangun model prediksi menggunakan algoritma Extreme Gradient Boosting (XGBoost).

6. Variabel Penelitian

Variabel penelitian merupakan atribut atau informasi yang digunakan dalam pro ses analisis dan pembangunan model prediksi. Dalam penelitian Prediksi Target Pendapatan Pajak Daerah di Kabupaten Sumbawa Menggunakan Algoritma Extreme Gradient Boosting (XGBoost), dataset yang digunakan terdiri dari 4 variabel utama, yaitu sebagai berikut:

Tabel 2. Variabel Dataset

No	Nama Variabel	Keterangan	Tipe data
1	Tahun	Menunjukkan periode waktu data pendapatan pajak	Integer
2	Jenis Pajak	Jenis pajak daerah seperti Pajak Hotel, Pajak Restoran, Pajak Reklame, Pajak PBB-P2, dan lain-lain	String
3	Target Pajak	Nilai target pendapatan pajak daerah yang telah ditetapkan oleh Bapenda per tahun	Float
4	Realisasi Pajak	Nilai aktual penerimaan pajak daerah yang tercapai pada tahun berjalan	Float

B. Metode Pengembangan Sistem

Penelitian ini menerapkan kerangka kerja data mining Cross Industry Standard Process for Data Mining (CRISP-DM), yang terdiri atas enam tahapan utama[13], yaitu:

1. Fase Pemahaman Bisnis (*Business Understanding*)

Pada fase ini, fokus utamanya adalah memahami tujuan dan kebutuhan bisnis secara menyeluruh, menilai kondisi yang ada, serta merumuskan sasaran dari proses *data mining*. Dalam konteks penelitian ini, dilakukan identifikasi terhadap permasalahan penelitian yang mencakup latar belakang, rumusan masalah, tujuan penelitian, serta batasan penelitian. Proses ini bertujuan agar arah penelitian menjadi jelas dan terarah, khususnya dalam upaya membangun model prediksi target pendapatan pajak daerah di Kabupaten Sumbawa.

2. Fase Pemahaman Data (*Data Understanding*)

Fase ini meliputi kegiatan pengumpulan data awal, penjabaran karakteristik data, eksplorasi data, serta pemeriksaan kualitas data. Pada tahap ini dilakukan proses pengumpulan dan ekstraksi data historis pajak daerah yang kemudian dianalisis untuk mengetahui struktur data, jenis variabel, jumlah data, serta kemungkinan adanya nilai kosong (*missing value*) atau anomali. Tahap ini bertujuan agar setiap atribut atau variabel dalam dataset dapat dipahami secara menyeluruh sebelum digunakan dalam proses analisis lebih lanjut.

3. Fase Pengolahan Data (*Data Preprocessing*)

Pada tahap ini, data disiapkan agar siap digunakan dalam pembangunan model *machine learning*. Proses yang dilakukan

meliputi pembersihan data (*data cleaning*), yaitu mengatasi nilai kosong atau data yang tidak konsisten; transformasi data (*data transformation*), yaitu mengubah data ke dalam format yang sesuai dengan kebutuhan model; serta reduksi data (*data reduction*) untuk menyederhanakan data tanpa menghilangkan informasi penting. Tahap ini memiliki peran yang sangat penting karena kualitas data yang baik akan sangat mempengaruhi performa model yang dihasilkan.

4. Fase Permodelan (*Modelling*)

Pada tahap ini dilakukan pembangunan model prediksi menggunakan algoritma *Extreme Gradient Boosting (XGBoost)*. Algoritma ini merupakan metode berbasis *ensemble learning* yang menggunakan teknik *boosting* untuk meningkatkan akurasi prediksi dengan cara membangun beberapa model secara bertahap[13]. Setiap model baru berfungsi untuk memperbaiki kesalahan dari model sebelumnya melalui optimasi *gradientdescent*.

Dalam proses pemodelan, beberapa parameter model seperti *learning rate*, *max depth*, dan *n_estimators* digunakan untuk mengatur proses pembelajaran model. Selanjutnya dilakukan tuning parameter (*hyperparameter tuning*) untuk menentukan kombinasi parameter terbaik sehingga diperoleh keseimbangan antara *bias* dan *variansi*. Model yang telah dibangun kemudian dilatih menggunakan data historis untuk menghasilkan prediksi target pendapatan pajak daerah.

5. Fase Evaluasi (*Evaluation*)

Fase ini dilakukan setelah model selesai dibangun dengan tujuan memastikan bahwa model memiliki kinerja dan kualitas yang baik dari sudut pandang analisis data. Dalam penelitian ini, pengujian performa model dilakukan menggunakan dua metrik evaluasi utama, yaitu *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan R-*Squared* (R^2), guna mengukur tingkat kesalahan dan akurasi hasil prediksi yang dihasilkan oleh model.

6. Fase Penyebaran (*Deployment*)

Hasil analisis dan model yang telah dibangun disajikan melalui dashboard interaktif berbasis flask. Penyebaran ini memungkinkan visualisasi data, hasil prediksi, serta rekomendasi analitis yang dapat digunakan oleh pemangku kebijakan dalam mendukung pengambilan keputusan strategis.

IV. HASIL DAN PEMBAHASAN

A. *Fase Pemahaman Bisnis (Business Understanding*)

Fase pemahaman bisnis adalah tahapan krusial yang mengedepankan pemahaman mendalam terhadap tujuan penelitian dari perspektif bisnis. Fase ini terbagi menjadi beberapa komponen utama, yaitu identifikasi masalah bisnis, penetapan tujuan bisnis dan *data mining*, serta penentuan metrik keberhasilan untuk penelitian ini.

1. Identifikasi Masalah

Dinamika Target pendapatan pajak daerah memiliki peran strategis dalam mendukung perencanaan anggaran belanja daerah di Kabupaten Sumbawa serta menjadi dasar kinerja Badan Pendapatan Daerah (Bapenda) Sumbawa. Namun, selama ini penentuan target masih dilakukan secara konvensional, yaitu menggunakan microsoft excel atau perhitungan sederhana seperti rata-rata dan *margin error*. Oleh karena itu, penelitian ini menganalisis data historis target dan realisasi pajak daerah serta mengembangkan model prediksi berbasis algoritma XGBoost guna membantu pemerintah dalam menetapkan target pajak yang lebih realistis, akurat, dan terukur.

2. Tujuan

Tujuan utama penelitian ini adalah untuk memprediksi target pendapatan pajak daerah dengan mengembangkan model *machine learning*. Model ini tidak hanya akan memprediksi target, tetapi juga akan mendukung pengambilan keputusan yang relevan berdasarkan hasil analisis prediksi yang telah dilakukan. Selain itu, model diimplementasikan menggunakan *framework*

flask sehingga proses prediksi dapat dilakukan secara *real-time*.

B. Fase Pemahaman Data (Data Understanding)

Fase ini berfokus pada pengumpulan serta pemahaman karakteristik data pajak yang akan digunakan dalam proses pemodelan. Tahapan ini meliputi proses pengumpulan data target dan realisasi pajak daerah, penjelasan struktur dan atribut data, eksplorasi pola dan tren data pajak, serta pemeriksaan kualitas data untuk memastikan bahwa data tersebut layak digunakan pada tahap pemodelan berikutnya.

1. Pengumpulan data

Tahap pengumpulan data dilakukan dengan mengumpulkan data historis terkait target dan realisasi pendapatan pajak daerah Kabupaten Sumbawa. Data diperoleh melalui dokumen laporan resmi pemerintah daerah, data statistik instansi terkait, serta sumber pendukung lain yang relevan. Proses ini memastikan bahwa seluruh jenis pajak, seperti pajak hotel, restoran, hiburan, reklame, parkir, dan pajak daerah lainnya, tercatat dengan lengkap dan konsisten untuk periode analisis yang dibutuhkan. Pengumpulan data juga diperkuat dengan studi literatur dan wawancara untuk memahami konteks dan faktor-faktor yang memengaruhi pendapatan pajak. Data yang dikumpulkan meliputi aspek target pendapatan dan realisasi penddapatan pajak daerah

2. Deskripsi Data

Ruang lingkup analisis pada penelitian ini difokuskan pada perkembangan target serta realisasi pendapatan pajak daerah di Kabupaten Sumbawa, dengan tujuan untuk memperoleh gambaran menyeluruh mengenai pola perubahan penerimaan, kontribusi tiap jenis pajak, dan kecenderungan pertumbuhan dari tahun ke tahun. Seluruh data terkait target dan realisasi pajak kemudian digabungkan dalam satu basis data sebagai bagian dari eksplorasi data, sehingga evaluasi terhadap tren historis, variasi nilai, serta hubungan antar variabel dapat dilakukan secara lebih efektif. Pemeriksaan kualitas data dilakukan secara menyeluruh untuk memastikan bahwa dataset yang digunakan valid, layak, dan representatif sebagai dasar dalam proses pengembangan model prediksi menggunakan algoritma XGBoost.

Tabel 3. Karakteristik Dataset Pendapatan Pajak Daerah

	Total Baris	Total Kolom	Kolom Tipe Data	Missing Value per Kolom
0	376	6	{'No': int64, 'Tahun': int64, 'Uraian Pajak': ...	{'No': 0, 'Tahun': 0, 'Uraian Pajak': 0, 'Jeni...

Tabel tersebut menunjukkan karakteristik dasar dataset yang terdiri dari 376 baris dan 6 kolom dengan berbagai tipe data, seperti *integer* pada kolom Tahun, *object* pada Uraian dan Jenis Pajak, serta *float* pada kolom numerik seperti Target dan Realisasi. Terdapat beberapa *missing value*, pada kolom Jenis Pajak yang memiliki 10 data kosong. Secara umum, tabel ini memberikan gambaran mengenai struktur, tipe data, dan kualitas dataset melalui identifikasi nilai yang hilang.

Tabel 4. Statistik Umum Target Pajak per Jenis Pajak

	Jenis Pajak	count	mean	median	min	max	std
0	Bea Perolehan Hak Atas Tanah dan Bangunan(BPHTB)	23	4.583174e+09	6.400000e+09	6.000000e+08	8.500000e+09	3.215964e+09
1	Pajak Hotel	2	1.680000e+09	1.680000e+09	9.000000e+07	3.270000e+09	2.248600e+09
2	Pajak Air Tanah	13	3.621527e+08	3.664500e+08	3.000000e+08	3.847725e+08	3.047762e+07
3	Pajak Bumi dan Bangunan Perdesaan dan Perkotaan	13	6.700000e+09	6.700000e+09	6.700000e+09	6.700000e+09	0.000000e+00
4	Pajak Hiburan	10	5.130391e+08	4.000000e+08	3.000000e+07	2.017391e+09	5.413275e+08
5	Pajak Hotel	17	1.421530e+09	2.280000e+09	0.000000e+00	2.912925e+09	1.311690e+09
6	Pajak Mineral bukan Logam dan Batuan	23	6.578101e+09	1.441611e+09	0.000000e+00	3.680294e+10	1.210354e+10
7	Pajak Parkir	10	4.866496e+08	4.900861e+08	4.725450e+08	4.961722e+08	1.087435e+07
8	Pajak Penerang Jalan	20	1.046710e+10	1.546050e+10	0.000000e+00	1.960000e+10	8.885356e+09
9	Pajak Prkir	2	4.300040e+08	4.300040e+08	4.300040e+08	4.300040e+08	0.000000e+00
10	Pajak Reklame	14	1.134931e+09	1.279558e+09	3.000000e+07	1.320060e+09	3.464212e+08
11	Pajak Restoran	20	2.191280e+09	2.600000e+09	0.000000e+00	4.281000e+09	1.340926e+09
12	Pajak Sarang Burung Walet	13	4.138462e+07	4.200000e+07	4.000000e+07	4.300000e+07	1.192928e+06

Tabel tersebut menunjukkan statistik deskriptif target pajak untuk setiap jenis pajak daerah di Kabupaten Sumbawa. BPHTB memiliki rata-rata sekitar Rp 4,58 miliar dengan standar deviasi tinggi, menandakan fluktuasi target yang besar antar tahun. PBB-P2 memiliki standar deviasi 0 sehingga targetnya selalu sama setiap tahun. Pajak Mineral Bukan Logam dan Batuan serta Pajak Penerangan Jalan menunjukkan variasi yang sangat besar, terlihat dari standar deviasi dan rentang nilai yang lebar. Sebaliknya, Pajak Air Tanah, Pajak Parkir, dan Pajak Sarang Burung Walet memiliki variasi rendah, menunjukkan bahwa target pajaknya relatif stabil sepanjang periode pengamatan.

Tabel 5. Statistik Umum Realisasi Pajak per Jenis Pajak

	Jenis Pajak	count	mean	median	min	max	std
0	Bea Perolehan Hak Atas Tanah dan Bangunan(BPHTB)	23	5.080967e+09	3.850153e+09	24831250.0	1.407324e+10	4.374321e+09
1	Pajak Hotel	2	2.792842e+07	2.792842e+07	16807500.0	3.904933e+07	1.572735e+07
2	Pajak Air Tanah	13	3.272234e+08	3.470995e+08	58398256.0	4.473740e+08	1.305456e+08
3	Pajak Bumi dan Bangunan Perdesaan dan Perkotaan	13	3.589178e+09	3.906167e+09	282349316.0	5.191653e+09	1.692424e+09
4	Pajak Hiburan	10	4.887902e+08	3.847220e+08	4758950.0	1.046856e+09	4.228658e+08
5	Pajak Hotel	17	1.863499e+09	1.892360e+09	120423300.0	3.799743e+09	1.571238e+09
6	Pajak Mineral bukan Logam dan Batuan	23	9.996062e+08	9.031892e+08	0.0	2.733151e+09	9.151847e+08
7	Pajak Parkir	10	5.016785e+08	5.296439e+08	398912020.0	5.560547e+08	6.073550e+07
8	Pajak Penerang Jalan	20	9.338474e+09	8.424320e+09	0.0	2.165696e+10	9.393237e+09
9	Pajak Prkir	2	3.925080e+07	3.925080e+07	39250800.0	3.925080e+07	0.000000e+00
10	Pajak Reklame	14	1.036582e+09	1.289255e+09	4758950.0	1.388423e+09	4.730726e+08
11	Pajak Restoran	20	2.749638e+09	2.625630e+09	0.0	6.578194e+09	1.986772e+09
12	Pajak Sarang Burung Walet	13	3.288684e+07	3.611980e+07	20650000.0	4.000000e+07	7.399245e+06

Tabel tersebut menyajikan statistik umum realisasi pajak untuk setiap jenis pajak daerah di Kabupaten Sumbawa. BPHTB memiliki rata-rata tertinggi dengan variasi besar, terlihat dari standar deviasi yang sangat tinggi. Pajak Penerangan Jalan juga menunjukkan fluktuasi yang signifikan, ditandai oleh rentang nilai minimum hingga maksimum yang sangat lebar. Sebaliknya, Pajak Parkir, Pajak Air Tanah, dan Pajak Sarang Burung Walet memiliki standar deviasi yang rendah, menandakan bahwa realisasinya relatif stabil dari tahun ke tahun. Beberapa jenis pajak, seperti PBB-P2 dan Pajak Parkir (baris tertentu), bahkan memiliki variasi sangat kecil hingga mendekati konstan.

Sementara itu, Pajak Reklame, Restoran, dan Mineral Bukan Logam dan Batuan menunjukkan variasi sedang dengan perubahan target atau realisasi yang masih terlihat antarperiode.

C. Fase Pengolahan Data (Data Preprocessing)

1. Pembersihan data

Tahap pembersihan dilakukan untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik valid, dan layak dijadikan fondasi pemodelan sebelum masuk ke proses pelatihan model. Pada penelitian ini, pembersihan data difokuskan pada konsistensi tipe data, kelengkapan nilai, serta validitas setiap variabel yang berperan dalam pemodelan.

Beberapa variabel numerik seperti tahun, realisasi, dan target masih ditemukan dalam format string atau mengandung karakter non-numerik. Kondisi tersebut memicu munculnya *missing value* ketika nilai yang tidak valid dikonversi menjadi NaN. Untuk mengatasi hal ini, dilakukan proses konversi tipe data menggunakan *pd.to_numeric* serta fungsi *to_numeric_clean*(), yang dirancang untuk membersihkan karakter tidak valid dan mengubah nilai yang tidak dapat diparsing menjadi NaN. Dengan demikian, seluruh nilai numerik berada dalam format yang sesuai dan siap diproses lebih lanjut.

Setelah proses konversi dilakukan, *missing value* yang muncul pada kolom-kolom kunci seperti tahun, realisasi, target, dan jenis_pajak ditangani dengan pendekatan data *removal* menggunakan *dropna*(). Baris yang memiliki nilai kosong dihapus untuk memastikan bahwa hanya data yang lengkap, valid, dan bebas dari *missing value* yang digunakan dalam pelatihan model. Langkah ini penting untuk menghindari bias model serta menjaga stabilitas dan keandalan hasil prediksi.

2. *Feature Engineering*

Feature engineering merupakan tahap penting dalam pengembangan model *machine learning*, untuk mengubah data mentah menjadi fitur yang lebih representatif, terstruktur, dan informatif. *Feature engineering* pada penelitian ini dilakukan untuk memperkaya informasi dalam dataset sehingga model mampu menangkap pola historis, karakteristik musiman, serta hubungan antarvariabel yang mungkin tidak terlihat secara langsung dari data awal.

Pada penelitian ini, variabel kategorikal jenis_pajak dikonversi ke bentuk numerik menggunakan *LabelEncoder*, yang kemudian menghasilkan fitur baru bernama jenis_pajak_*encoded*. Transformasi ini sangat penting mengingat sebagian besar algoritma machine learning, termasuk XGBoost, tidak dapat mengolah data kategorikal secara langsung. Dengan adanya pengkodean tersebut, informasi kategoris tetap dapat dipertahankan tanpa kehilangan karakteristik tiap kategori pajak.

3. Transformasi data

Transformasi data dilakukan menggunakan metode *log-transform* untuk menyeimbangkan skala antar fitur. Proses ini mencegah dominasi fitur tertentu dalam pelatihan model dan

dilakukan dengan fungsi *StandardScaler* dari pustaka *Scikit-Learn*. Hasil transformasi membantu meningkatkan stabilitas dan konsistensi performa model XGBoost.

4. *Split* data

Split data merupakan tahap penting dalam proses pra-pemrosesan dan pengembangan model *machine learning*. Dalam penelitian ini, dataset dibagi menggunakan rasio standar 80:20, di mana 80% data dialokasikan sebagai *Training set* untuk melatih model, sedangkan 20% sisanya dijadikan *testing set* untuk mengevaluasi kinerja prediksi model secara objektif. Pembagian ini dilakukan menggunakan fungsi *train_test_split* dari *scikit-learn* dengan parameter *test_size*=0.2, yang menunjukkan proporsi data pengujian, serta *random_state*=42 untuk memastikan bahwa proses pemisahan bersifat acak namun tetap konsisten dan dapat direproduksi. Implementasi pembagian tersebut ditunjukkan melalui kode *X_train, X_test, y_train, y_test = train_test_split(X, y_log, test_size=0.2, random_state=42)*, yang menghasilkan empat bagian data: fitur pelatihan (*X_train*), fitur pengujian (*X_test*), target pelatihan (*y_train*), dan target pengujian (*y_test*). Dengan proses ini, model dapat dievaluasi secara lebih akurat karena pengujian dilakukan pada data yang tidak digunakan selama pelatihan.

D. Fase Permodelan (Modelling)

Fase ini menjelaskan proses pembangunan model *Extreme Gradient Boosting* (XGBoost), mulai dari arsitektur algoritma hingga konfigurasi model yang digunakan. Setelah tahap pra-pemrosesan data selesai, model prediksi dikembangkan menggunakan algoritma XGBoost Regressor dengan dukungan pustaka python pada ekosistem *Scikit-learn*. Model awal dikonfigurasi menggunakan parameter inti *objective='reg:squarederror'*,n_*estimators*=500, *learning_rate*=0.05, serta *max_depth*=4 sebagai batas kedalaman pohon untuk menjaga stabilitas proses pembelajaran dan mencegah overfitting.

Parameter *objective='reg*:*squarederror*' berfungsi menetapkan bahwa model bekerja pada kasus regresi dengan tujuan meminimalkan error kuadrat antara nilai prediksi dan nilai aktual, sehingga model fokus pada akurasi nilai kontinu. Selanjutnya, n_*estimators*=500 menunjukkan jumlah pohon (*decision tree*) yang dibangun dalam proses boosting, dimana semakin banyak pohon memberi kemampuan model untuk mempelajari pola data lebih dalam. Sementara itu, *learning_rate*=0.05 mengontrol seberapa besar kontribusi setiap pohon dalam update model, sehingga proses pembelajaran berlangsung lebih halus dan mengurangi risiko *overfitting*. Proses pelatihan dilakukan dengan memanggil fungsi model.fit (*X_train, y_train*) sehingga model mampu mempelajari pola historis data dan menghasilkan prediksi target pendapatan pajak daerah secara lebih akurat.

E. Fase Evaluasi

Evaluasi model dilakukan menggunakan data uji untuk menilai akurasi prediksi target pendapatan pajak daerah. Kinerja model diukur melalui beberapa metrik, yaitu RMSE untuk melihat besarnya kesalahan prediksi, MAPE untuk mengukur persentase error, serta R-*Squared* (R^2) untuk menilai seberapa baik model menjelaskan variabilitas data. Nilai evaluasi model Sebagai berikut:

Tabel 6. Nilai Evaluasi Model

Metrik Evaluasi	Nilai Hasil Model
RMSE	777,824,418.79
MAPE	5,14%
R^2	0,98

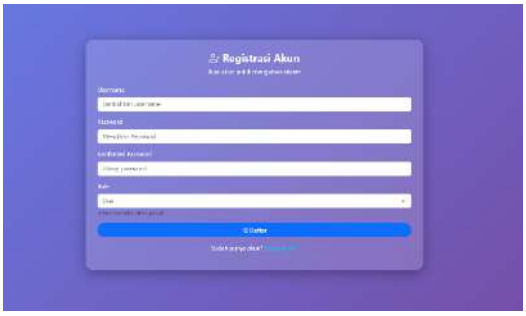
Berdasarkan tabel tersebut hasil evaluasi menunjukkan bahwa model prediksi bekerja sangat baik dalam memperkirakan target pendapatan pajak daerah. Nilai RMSE sebesar 777,824,418.79 atau kisaran 777 juta rupiah menunjukkan bahwa rata-rata selisih prediksi terhadap nilai aktual masih wajar karena target berada pada skala miliaran rupiah. Nilai MAPE 5,14% menegaskan bahwa tingkat kesalahan prediksi sangat rendah. Sementara itu, R^2 sebesar 0,98 menunjukkan bahwa model

mampu menjelaskan 98% variasi data. Secara keseluruhan, model ini stabil, akurat, dan layak digunakan untuk mendukung penetapan target pendapatan pajak daerah.

F. Fase Penyebaran (Deployment)

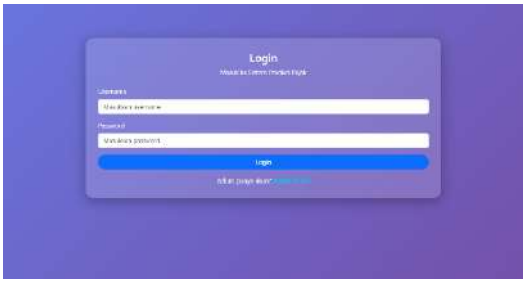
Setelah proses prediksi dan evaluasi model dilakukan, langkah berikutnya adalah menampilkan hasil prediksi dalam sebuah antarmuka aplikasi berbasis *web*. Implementasi sistem dilakukan menggunakan framework *flask*, yang berfungsi sebagai penghubung antara model machine learning dengan pengguna. Melalui *flask*, model yang telah dilatih dapat menerima input baru dari user, memprosesnya, kemudian menghasilkan nilai prediksi yang akan ditampilkan langsung ke halaman *web*. Aplikasi berbasis *web* ini dibangun dengan empat menu utama yaitu sebagai berikut:

1. Register
   Halaman Register digunakan oleh pengguna baru untuk membuat akun sebagai syarat agar dapat masuk dan menggunakan sistem.



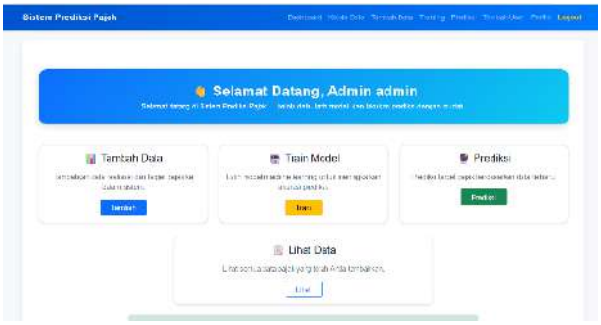
Gambar 3. Halaman Register

2. Login
   Halaman Login merupakan halaman yang digunakan pengguna untuk mengakses sistem dengan memasukkan akun yang telah terdaftar.



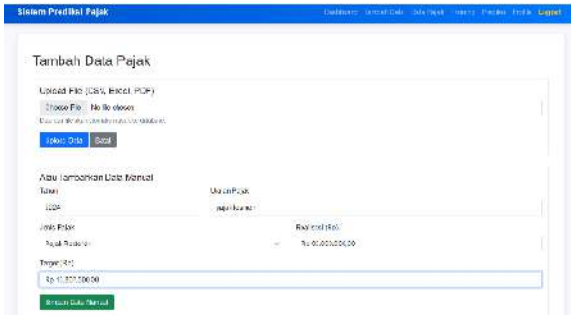
Gambar 4. Halaman Login

3. Dashboard
   Halaman dashboard berfungsi sebagai pusat kontrol sistem yang menyediakan berbagai fitur utama, seperti tambah data, train model, prediksi, dan lihat data. Secara keseluruhan, dashboard memudahkan pengelolaan data serta proses prediksi dalam satu tampilan yang terintegrasi.



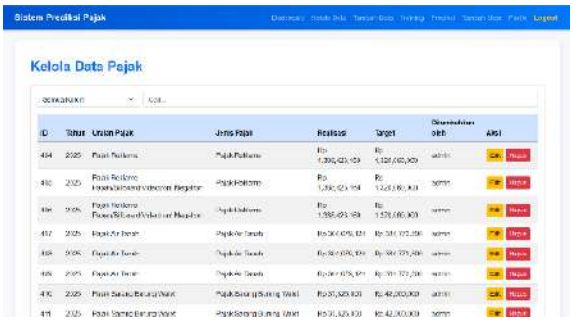
Gambar 3. Halaman Dashboard

4. Tambah Data
   Menu Tambah Data berfungsi untuk menambahkan data baru ke dalam database yang akan digunakan sebagai bahan pelatihan maupun pembaruan dataset.



Gambar 4. Halaman Tambah Data

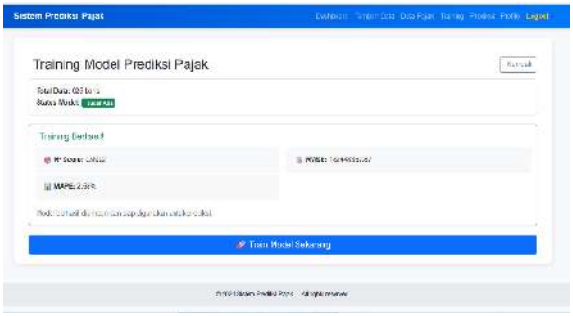
5. Kelola data
   Halaman kelola data hanya dapat diakses oleh admin, sehingga hanya admin yang memiliki wewenang untuk mengedit dan menghapus data yang telah ditambahkan.



Gambar 5. Halaman Kelola Data

6. *Training* Model

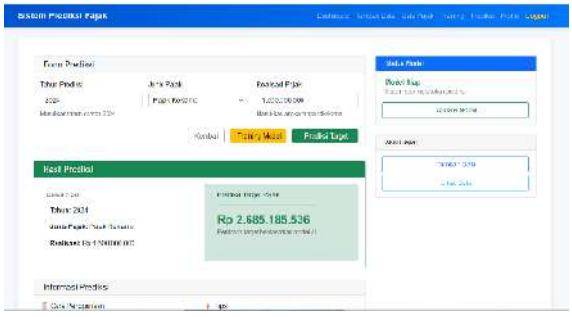
Menu *Training* Model menyediakan fitur untuk melatih ulang model apabila terdapat data terbaru atau memperbarui model prediksi menggunakan data terbaru secara langsung melalui antarmuka sistem.



Gambar 6. Halaman *Training* Model

7. Prediksi

Halaman Prediksi digunakan sebagai form yang memungkinkan pengguna memasukkan nilai realisasi pajak untuk mendapatkan hasil prediksi target pendapatan pajak secara otomatis.



Gambar 7. Halaman Prediksi

V. KESIMPULAN

Berdasarkan hasil penelitian, penerapan algoritma *Extreme Gradient Boosting* (XGBoost) terbukti efektif dalam memprediksi target pendapatan pajak daerah di Kabupaten Sumbawa. Model XGBoost mampu mengolah variabel tahun, jenis pajak, dan realisasi pajak sehingga menghasilkan pola prediksi yang akurat dan stabil. Evaluasi menggunakan metrik RMSE, MAPE, dan R^2 menunjukkan bahwa model memiliki performa prediksi yang sangat baik, dengan nilai RMSE sebesar 777 juta, MAPE sebesar 5,14%, dan R^2 sebesar 98%. Hasil tersebut menegaskan bahwa model mampu menjelaskan sebagian besar variasi data serta menghasilkan deviasi prediksi yang relatif kecil terhadap nilai aktual.

Selain itu, prototype prediksi yang diintegrasikan ke dalam dashboard *Flask* memberikan kemudahan bagi pemerintah daerah dalam melakukan train model dan memprediksi target pendapatan pajak secara real time. Sistem yang dikembangkan menyediakan fitur pengelolaan data, proses pelatihan model (*training*), serta proses memprediksi target pajak secara real time. Dengan adanya sistem ini, pemerintah daerah dapat memperoleh informasi prediktif yang lebih akurat dan berbasis data sehingga dapat membantu dalam proses penetapan target pajak serta perencanaan keuangan daerah secara lebih terukur.

Namun demikian, penelitian ini masih memiliki beberapa keterbatasan. Data yang digunakan dalam penelitian ini terbatas pada periode 2021-2025 dan hanya menggunakan beberapa variabel utama, yaitu tahun, jenis pajak, dan realisasi pajak. Selain itu, penelitian ini hanya menerapkan satu algoritma *machine learning*, yaitu XGBoost, sehingga belum dilakukan perbandingan dengan algoritma lain yang mungkin memiliki performa yang berbeda.

Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset dengan periode waktu yang lebih panjang serta menambahkan variabel lain yang berpotensi memengaruhi penerimaan pajak daerah. Selain itu, penelitian selanjutnya juga dapat melakukan perbandingan dengan berbagai algoritma *machine learning* lainnya atau

mengembangkan metode *ensemble* untuk meningkatkan akurasi prediksi. Pengembangan sistem juga dapat ditingkatkan dengan menambahkan fitur visualisasi data yang lebih interaktif serta integrasi dengan sistem informasi pemerintah daerah.

DAFTAR PUSTAKA

[1] H. Yasser and T. D. Widajantie, "Pengaruh Pajak Daerah dan Retribusi Daerah Terhadap Pendapatan Asli Daerah Provinsi Jatim," *JIMEA | J. Ilm. MEA (Manajemen, Ekon. dan Akuntansi)*, vol. 6, no. 1, pp. 611–619, 2022.

[2] Mahfudh, H. Saleh, and M. Y. Saleh, *Analisis Peningkatan Pendapatan Asli Daerah*, vol. 1, no. 1. 2022.

[3] J. Martaviona and S. Nurhalimah, "Peran pajak daerah dalam pembangunan ekonomi lokal," vol. 3, no. 1, pp. 265–277, 2025.

[4] Munaldi, *Algoritma Random Forest Dan Xgboost Menggunakan Python*. PT. GANESHA KREASI SEMESTA, 2025.

[5] F. Puteri Marchanda Izzati, "Implementasi Algoritma XGBoost Untuk Prediksi Capaian Bulanan Pendapatan Daerah Kota Bandung," *J. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, pp. 104–111, 2025.

[6] S. Hasibuan, "Analisis Clustering Pendapatan Pajak dan Retribusi Daerah di Kabupaten Sumbawa Dengan Model KDD Menggunakan Algoritma K-Means," *J. Inform. J. Pengemb. IT*, no. xx, pp. 1–10, 2024.

[7] M. E. Jasman Pardede, "Prediksi Jumlah Target dan Realisasi Wajib Pajak Atas PBB – P2 Menggunakan Multi Regression, Regression Lasso, dan PCR," *J. Edukasi dan Penelit. Inform.*, vol. 10, no. 1, p. 1, 2024, doi: 10.26418/jp.v10i1.68890.

[8] T. S. Muhammad Hafid Fikrianto, "Indonesian Journal of Law and Economics Review," *Forecast. Sidoarjo Local Tax Revenue Using Exponential Smoothing Model.*, vol. 20, no. 4, 2025, doi: 10.21070/ijler.v20i4.1458.

[9] M. Irham and A. Saputra, "Forecasting Value-Added Tax (VAT) Revenue Using Autoregressive Integrated Moving Average (ARIMA) Box-Jenkins Method," *J. Kaji. Ilm. Perpajak. Indones.*, vol. 4, no. 2, pp. 205–218, 2023, doi: 10.52869/st.v4i2.568.

[10] A. P. P. Saragih, E. Irawan, and M. Safii, "Optimalisasi Metode Conjugate Gradient Polak Rebiere Untuk Memprediksi Target PBB Kecamatan Dolok Merawan," *Semin. Nas. Inform.*, vol. 4, no. 3, pp. 241–248, 2023, [Online]. Available: https://www.ejournal.pelitaindonesia.ac.id/ojs32/index.php/SENATIKA/article/view/3701

[11] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *Nuansa Inform.*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.

[12] F. Sulianta, "Buku Dasar Data Mining from A to Z," in *Universitas Widyatama*, no. January, 2023, pp. 15–15.

[13] I. Daqiqil, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. 2021.