

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah mentransformasi industri pendidikan, dimana platform pembelajaran online berbasis keterampilan mengalami pertumbuhan yang sangat pesat dalam beberapa tahun terakhir. Platform seperti bootcamp dan kelas online tidak hanya menyediakan konten edukatif, tetapi juga menghasilkan data besar yang dapat dimanfaatkan untuk mendukung pengambilan keputusan berbasis data [1], [2]. Namun, pertumbuhan yang cepat ini menyebabkan tantangan operasional yang harus diatasi, khususnya dalam memprediksi fluktuasi jumlah peserta dan pendapatan bulanan yang sangat dinamis dan sulit diprediksi dengan metode konvensional.

Ketidakpastian pertumbuhan peserta dan pendapatan merupakan masalah kritis yang berdampak langsung pada stabilitas dan keberlanjutan bisnis platform pembelajaran online. Data historis menunjukkan adanya fluktuasi yang signifikan, dimana penurunan jumlah peserta dapat mencapai 20% setiap kuartal, menyebabkan ketidakstabilan pendapatan dan kesulitan dalam perencanaan keuangan jangka panjang. Pendapatan bulanan juga menunjukkan variasi yang besar, dengan penurunan hingga 30% pada beberapa bulan yang lebih sepi. Ketidakmampuan memprediksi fluktuasi ini menghambat capacity planning untuk infrastruktur, budget allocation untuk pengembangan konten dan marketing, serta strategic planning untuk ekspansi bisnis. Oleh karena itu, kemampuan melakukan forecasting yang akurat menjadi kebutuhan mendesak untuk menjaga efisiensi operasional dan keberlanjutan platform pembelajaran berbasis keterampilan.

Untuk mengatasi tantangan prediksi tersebut, berbagai penelitian menunjukkan bahwa pendekatan berbasis machine learning mampu memberikan hasil yang lebih akurat dibandingkan metode statistik tradisional. Algoritma seperti Random Forest dan Extreme Gradient Boosting terbukti efektif dalam memproyeksikan revenue dari data historis yang besar dan variative [3][4]. Voting Regressor mampu

meningkatkan akurasi prediksi pendapatan tahunan dan partisipasi peserta, meminimalkan kesalahan prediksi, serta memberikan informasi mengenai program yang paling diminati [5], [6]. Pendekatan machine learning ini memungkinkan platform memahami pola pembelian peserta, memaksimalkan profit, dan mengidentifikasi program unggulan [7], [8]. Model ensemble seperti Random Forest dan Extreme Gradient Boosting sering memberikan hasil lebih baik dibandingkan model regresi tunggal, dimana Extreme Gradient Boosting mampu menangani data tidak terstruktur dan memberikan prediksi yang stabil untuk pendapatan dan partisipasi peserta [3], [4].

Selain machine learning, pendekatan deep learning juga menunjukkan keunggulan dalam menangani data time-series yang sangat bervariasi. Model Long Short-Term Memory digunakan untuk memprediksi tingkat partisipasi peserta secara akurat, sementara Deep Neural Networks mampu memproses data yang lebih kompleks, menghasilkan proyeksi pendapatan dan partisipasi yang lebih stabil [9], [10]. Penelitian menunjukkan bahwa Long Short-Term Memory dan ARIMA dapat menganalisis fluktuasi data historis dengan baik untuk time-series forecasting [10], [11], [12].

Namun, performa model machine learning dan deep learning sangat bergantung pada konfigurasi hyperparameter yang tepat. Optimasi hyperparameter merupakan langkah penting untuk meningkatkan akurasi model prediksi. Teknik seperti Optuna, Hyperband, dan Bayesian Optimization terbukti meningkatkan performa model machine learning dan deep learning, memungkinkan prediksi lebih cepat, [3], [13], [14], [15]. Kombinasi machine learning dan deep learning meningkatkan kemampuan platform untuk memprediksi kinerja keuangan dan engagement peserta secara simultan.

Dalam edukasi teknologi Indonesia, Labskill merupakan platform pembelajaran online berbasis keterampilan yang menyediakan berbagai kursus dan pelatihan untuk pengembangan kompetensi digital. Sebagai platform yang relatif baru dengan periode operasional 24 bulan dari Januari 2023 hingga Desember 2024, Labskill menghadapi tantangan yang sama dengan platform edukasi online pada

umumnya, yaitu fluktuasi signifikan dalam jumlah peserta dan pendapatan bulanan. Ketidakpastian ini menghambat kemampuan manajemen dalam melakukan capacity planning, budget allocation, dan strategic planning. Data historis Labskill menunjukkan variasi yang dipengaruhi oleh berbagai faktor seperti harga kursus, jumlah referral, dan tingkat completion rate pengguna. Pemanfaatan Business Intelligence dan visual analytics dapat membantu pengambilan keputusan berbasis data, sehingga strategi pemasaran, program pembelajaran, dan pengelolaan keuangan dapat dilakukan lebih efektif [1], [2]. Kompleksitas interaksi antar faktor ini memerlukan pendekatan forecasting yang mampu memodelkan relationship non-linear dan memberikan prediksi akurat untuk mendukung pengambilan keputusan berbasis data.

Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan dan membandingkan pendekatan machine learning dan deep learning untuk memprediksi pertumbuhan peserta dan pendapatan platform Labskill. Secara spesifik, penelitian ini menggunakan Extreme Gradient Boosting sebagai representasi algoritma tree-based machine learning dan Artificial Neural Network sebagai representasi deep learning. Pemilihan Extreme Gradient Boosting didasarkan pada temuan literatur yang menunjukkan superioritas Extreme Gradient Boosting dalam sales forecasting, dimana Extreme Gradient Boosting terbukti sebagai model paling akurat di antara berbagai algoritma yang diuji untuk prediksi revenue dan pertumbuhan [16] dengan karakteristik optimisasi kecepatan dan efisiensi memori yang ideal untuk data berdimensi tinggi dengan relationship kompleks [17]. Sementara itu, Artificial Neural Network dipilih sebagai pembanding untuk memvalidasi efektivitas pendekatan tree-based pada dataset terbatas, mengingat penelitian menunjukkan Artificial Neural Network efektif untuk prediksi pada kasus kompleks dengan dataset yang besar, mencapai akurasi sangat tinggi 0.37% [13], namun performanya sangat bergantung pada jumlah data training yang cukup untuk menghindari overfitting.

Kedua algoritma dikembangkan dalam dua versi: model baseline dengan hyperparameter default dan model optimized dengan hyperparameter hasil tuning

menggunakan Optuna untuk Extreme Gradient Boost dan Hyperband untuk Artificial Neural Network. Pemilihan Optuna untuk Extreme gradient Boosting didasarkan pada temuan bahwa Bayesian optimization efektif meningkatkan performa algoritma tree-based, dimana Bayesian optimization efektif meningkatkan performa algoritma tree-based, dimana dengan Bayesian Optimization mencapai R^2 0.85 mengungguli model baseline pada [10], [11], Sementara Hyperband dipilih untuk Artificial Neural Network karena efisiensinya dalam menangani search space hyperparameter yang kompleks pada deep learning melalui successive halving [18], [4]. Dataset penelitian mencakup 24 observasi bulanan dengan 8 features hasil feature engineering yang merepresentasikan karakteristik transaksi, pricing, dan behavioral patterns pengguna Labskill. Model terbaik dari komparasi ini kemudian diimplementasikan dalam aplikasi forecasting berbasis Streamlit untuk prediksi 6 bulan ke depan, memberikan tool praktis yang dapat digunakan langsung oleh stakeholder Labskill untuk decision making [19], [20].

Penelitian ini memberikan beberapa kontribusi penting. Pertama, penelitian ini menyediakan komparasi sistematis antara algoritma tree-based dan neural network pada forecasting platform edukasi online dengan dataset terbatas (24 observasi), dimana sebagian besar penelitian terdahulu fokus pada dataset yang lebih besar [9], [10]. Kedua, penelitian ini mengeksplorasi efektivitas strategi optimasi hyperparameter yang berbeda untuk setiap kategori algoritma, memberikan guidance praktis untuk pemilihan teknik tuning yang sesuai [13], [14], [15]. Ketiga, penelitian ini tidak hanya berkontribusi secara teoritis melalui validasi performa algoritma, tetapi juga memberikan implementasi praktis berupa aplikasi forecasting yang dapat langsung digunakan untuk capacity planning dan budget allocation [19], [20], [21]. Hasil penelitian ini diharapkan dapat membantu Labskill dalam memproyeksikan pendapatan dan pertumbuhan peserta dengan lebih akurat, sehingga meningkatkan efisiensi operasional dan merumuskan strategi bisnis yang lebih efektif [1], [2], [7], [8], [9]. Analisis berbasis data ini akan memungkinkan pengambilan keputusan lebih tepat, efisien, dan berbasis bukti, mendukung

keberlanjutan bisnis dan peningkatan kualitas layanan edukasi berbasis keterampilan [17].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini berupaya menjawab beberapa permasalahan utama berikut:

1. Bagaimana hasil memprediksi pendapatan bulanan LabSkill menggunakan data historis pelanggan dan program bootcamp?
2. Bagaimana hasil prediksi pertumbuhan peserta dengan kinerja Perusahaan platform LabSkill?
3. Bagaimana Hasil optimasi parameter model Machine Learning dan Deep Learning dengan optimasi hyperparameter berpengaruh terhadap tingkat error prediksi pendapatan dan pertumbuhan peserta?

1.3 Batasan Masalah

Agar penelitian tetap fokus dan realistik mengingat keterbatasan sumber daya perusahaan, penelitian ini membatasi diri pada:

1. Analisis hanya menggunakan model Machine Learning Extreme Gradient Boost dan Deep Learning Artificial Neural Network tanpa mengimplementasikan ensemble atau hybrid yang kompleks.
2. Optimasi hyperparameter dilakukan menggunakan Optuna dan Hyperband untuk meningkatkan performa model tanpa perlu membangun jaringan hybrid atau ensemble besar.
3. Data yang digunakan hanya mencakup periode 2023–2025 dan berfokus pada informasi pelanggan, program bootcamp, transaksi, serta engagement peserta.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang dapat membantu LabSkill dalam membuat estimasi pendapatan bulanan berdasarkan data historis pelanggan, program bootcamp, transaksi, serta

tingkat engagement peserta. Selain itu, penelitian ini bertujuan untuk memprediksi pertumbuhan peserta atau kemungkinan peningkatan partisipasi di platform LabSkill, sehingga perusahaan dapat merancang strategi retensi dan marketing yang lebih efektif.

Penelitian ini juga menganalisis pengaruh penggunaan model Machine Learning dan Deep Learning dengan optimasi hyperparameter terhadap akurasi prediksi, sehingga dapat memberikan panduan yang praktis bagi perusahaan yang memiliki keterbatasan sumber daya dalam membangun model ensemble atau hybrid yang kompleks. Dengan demikian, penelitian ini diharapkan dapat menghasilkan prediksi yang lebih akurat dan mendukung pengambilan keputusan berbasis data di LabSkill.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat praktis dan akademis bagi LabSkill. Secara praktis, hasil penelitian dapat membantu manajemen dalam merencanakan strategi keuangan dan program berdasarkan prediksi pendapatan tahunan yang lebih akurat. Selain itu, prediksi terkait pertumbuhan peserta dan tingkat partisipasi dapat digunakan untuk meningkatkan strategi retensi, engagement, dan efektivitas pemasaran, sehingga platform dapat menyesuaikan program sesuai kebutuhan dan preferensi peserta.

Dari sisi akademis, penelitian ini memberikan panduan praktis terkait pemanfaatan Machine Learning dan Deep Learning dengan optimasi hyperparameter sebagai alternatif yang efisien bagi perusahaan yang memiliki keterbatasan sumber daya, tanpa harus membangun model ensemble atau hybrid yang kompleks. Dengan demikian, penelitian ini tidak hanya mendukung pengambilan keputusan berbasis data, tetapi juga memberikan kontribusi pada pengembangan metode analitik untuk platform pembelajaran berbasis keterampilan.

1.5 Sistematika Penulisan

Sistematika penulisan berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, mulai dari pendahuluan hingga kesimpulan dan saran.

BAB I Bab ini berisi latar belakang, perumusan masalah, tujuan, ruang lingkup, dan batasan masalah yang digunakan serta sistematika penulisan.

BAB II Bab ini berisi teori-teori dasar yang digunakan sebagai landasan dalam proses pembuatan tugas akhir ini.

BAB III Bab ini menerangkan secara umum analisis dan perancangan independent data mart terkait penelitian yang dilakukan.

BAB IV Bab ini menerangkan hasil dan pembahasan analisis perancangan model machine learning dan deep learning terkait penelitian yang dilakukan.

BAB V Bab ini berisi tentang kesimpulan dan saran yang diperoleh dalam pembahasan penelitian ini untuk penyempurnaan dan pengembangan selanjutnya.

