

**PERANCANGAN *MACHINE LEARNING* PREDIKSI UNIT PELAKSANA TEKNIS
(UPT) YANG BERPOTENSI TEMUAN KERUGIAN KEUANGAN NEGARA
BERDASARKAN KINERJA KEUANGAN**
(Studi kasus di Kementerian Kementerian Perhubungan)

Eka C. S., ST, MAK, CFrA, CSA, CDSP
Alprisda WP,
Safira A, SAK

Abstraksi

This study describes the design of a machine learning model that predicts auditees with the potential for financial losses based on financial performance. The model can be used to support the selection of auditees for the annual supervision Internal Audit program of Internal Audit of the Ministry of Transportation. The machine learning algorithms used are a combination of logistic regression, random forest, and k-Nearest Neighbors algorithm. The final results of the assessment, demonstrate that the model was 81% accurate in predicting auditees with potential findings of state financial losses in 2021 which were published in the performance audit results report for 2022.

1. Latarbelakang

Salah satu perwujudan peran APIP yang efektif sebagaimana PP 60 Tahun 2008 tentang SPIP pada pasal 11 huruf a, **“sekurang-kurangnya harus memberikan keyakinan yang memadai atas ketaatan, kehematan, efisiensi, dan efektivitas pencapaian tujuan penyelenggaraan tugas dan fungsi Instansi Pemerintah”**. Dengan demikian, wajar apabila salah satu bentuk temuan kerugian negara yang diungkapkan oleh tim Auditor dipandang salah satu temuan strategis yang selalu diprioritaskan dalam pembahasan rapat pimpinan, oleh karena itu nilai Belanja Barang, Modal dan PNBPN menjadi salah satu komponen *risk factor* dalam penentuan UPT yang akan dikunjugi.

Inspektorat Jenderal Kementerian Perhubungan dalam pelaksanaan Program Kegiatan Pengawasan Tahunan (PKPT) untuk pemilihan UPT yang akan dilakukan pemeriksaan berdasarkan penilaian risiko yang berdasarkan kriteria-kriteria yaitu: 1)Jumlah pegawai; 2)Belanja Barang Operasional; 3)Belanja Barang Non-Operasional; 4)Modal; 5)PNBP; 6)Anggaran; 7)Beban Kerja; 8)Potensi PNBPN; 7)Penilaian Risiko; 8)Jangka Waktu Audit Sebelumnya; 9)Posisi Tindak Lanjut; 10)Lokasi geografis dan menurut survei peneliti terhadap pemilihan UPT dari tahun 2018-2021 diketahui bahwa rata-rata nilai persentase auditi yang terungkap dengan temuan kerugian negara sebesar 52%, sehingga memunculkan pertanyaan dari Peneliti apakah kriteria pemilihan UPT yang dipakai sudah cukup atau masih kurang.

Sebagaimana yang diuraikan sebelumnya bahwa Inspektorat Jenderal Kementerian Perhubungan belum mempertimbangkan analisa komponen laporan keuangan dalam salah satu kriteria pemilihan UPT. Pada tahun 2022, terdapat penelitian yang telah dilakukan oleh peneliti yang menunjukkan adanya pengaruh kinerja keuangan terhadap indikasi korupsi pada Kementerian Perhubungan, dimana indikasi korupsi adalah temuan kerugian negara sedangkan ukuran kinerja keuangan dimaksud terdiri dari nilai Likuiditas, Efisiensi Anggaran, Pertumbuhan Pendapatan dan Pertumbuhan Belanja.

Berdasarkan fenomena diatas dan seiring dengan perkembangan jaman revolusi industri 4.0 yaitu mengimplementasikan teknologi untuk mempermudah dalam pelaksanaan audit, sehingga mendorong kami untuk merancang sebuah pemograman *machine learning* yang dapat memprediksi UPT yang berpotensi terdapat temuan kerugian negara berdasarkan kinerja keuangan.

Tujuan penelitian adalah mengetahui keberhasilan *machine learning* dalam memprediksi UPT yang berpotensi terdapat temuan kerugian negara. Adapun algoritma *machine learning* yang akan

diujicobakan adalah algoritma *logistic regression*, *random forest* dan *k-Nearest Neighbors (KNN)* dimana ketiganya yang terbaik digunakan untuk *machine learning for fraud prediciton* (sumber: <https://sqream.com/blog/fraud-detection-machine-learning>)

2. Studi Pustaka

a. *Machine learning*

Machine learning adalah ilmu pengembangan algoritme dan model secara statistik yang digunakan sistem komputer untuk menjalankan tugas tanpa instruksi eksplisit, mengandalkan pola serta inferensi sebagai gantinya. Sistem komputer menggunakan algoritma *machine learning* untuk memproses data historis berjumlah besar dan mengidentifikasi pola data. Hal ini memungkinkannya untuk memprediksi hasil yang lebih akurat dari set data input yang diberikan. Misalnya, ilmuwan data dapat melatih aplikasi medis untuk mendiagnosis kanker dari gambar sinar-x dengan cara menyimpan jutaan gambar yang dipindai dan diagnosis yang sesuai. (sumber: <https://aws.amazon.com/id/what-is/machine-learning/>).

Gagasan utama di balik *machine learning* adalah hubungan matematis yang ada antara semua kombinasi data input dan output. Model *machine learning* tidak mengetahui hubungan ini sebelumnya, tetapi dapat menerka jika diberikan set data yang cukup. Hal ini berarti setiap algoritma *machine learning* dibangun di seputar fungsi matematika yang dapat dimodifikasi. Meskipun ini adalah pemahaman dasar, *machine learning* berfokus pada prinsip bahwa semua poin data kompleks dapat dihubungkan secara matematis oleh sistem komputer selama sistem komputer tersebut memiliki data dan daya komputasi yang cukup untuk memproses data tersebut. Oleh karena itu, keakuratan output tersebut secara langsung tergantung pada algoritma yang digunakan, kualitas dan kecukupan data input.

Machine learning membantu bisnis dengan cara mendorong pertumbuhan, membuka aliran pendapatan baru, dan menyelesaikan masalah-masalah yang menantang. Dilansir dari situs resmi Amazon, beberapa industri diketahui telah mengimplementasikan *machine learning*, yaitu: 1)Manufaktur; 2)Kesehatan; 3)Keuangan; 4)Ritel; dan 5)Media hiburan.

Hal-hal yang perlu dipertimbangkan dalam implementasi *machine learning* adalah: 1)kecukupan data set untuk pembelajaran *machine learning*; 2)untuk pemrosesan data yang besar memerlukan perangkat keras yang mumpuni; 3)Membutuhkan pemahaman yang cukup dalam pemrograman komputer dan interpretasi data dari ahli agar dapat dimengerti oleh publik.

b. Pengukuran Kinerja Keuangan

Pengukuran kinerja keuangan pada Sektor Privat maupun Publik merupakan variabel keuangan berupa rasio-rasio angka yang bersumber dari laporan keuangan entitas tersebut. Alasan-alasan tersebut digunakan untuk menganalisis kinerja keuangan. Rasio adalah panduan yang berguna untuk menilai situasi keuangan dan operasi perusahaan dan membandingkan hasil dengan tahun-tahun sebelumnya (Kasmir 2008). Di lingkungan pemerintah pusat, beberapa jenis rasio dapat dirumuskan berdasarkan data keuangan yang diperoleh dari laporan keuangan pemerintah pusat, yaitu:

1. Likuiditas, (Rasio Kas) menunjukkan kemampuan pemerintah untuk memenuhi kewajiban jangka pendeknya atau untuk melihat kemampuan pemerintah untuk mendanai kebutuhan. Terdapat penelitian sebelumnya yang menunjukkan nilai likuiditas berpengaruh negatif signifikan terhadap indikasi korupsi (Setyawan,2021) ;
2. Efisiensi anggaran, digunakan untuk mengukur tingkat penghematan anggaran pemerintah pusat. Tingkat efisiensi kegiatan pemerintah pusat dapat mempengaruhi kinerja keuangan pemerintah pusat dengan menunjukkan apakah pemerintah pusat menggunakan semua faktor produksinya secara efektif dan efisien. Terdapat penelitian sebelumnya yang menunjukkan nilai efisiensi anggaran berpengaruh negatif signifikan terhadap indikasi korupsi (Setyawan,2021);

3. Pertumbuhan pendapatan, analisis pertumbuhan pendapatan membantu menentukan capaian kinerja anggaran pemerintah pusat selama tahun anggaran yang bersangkutan. Terdapat penelitian sebelumnya yang menunjukkan nilai pertumbuhan pendapatan berpengaruh positif signifikan terhadap indikasi korupsi (Setyawan,2021);
4. Pertumbuhan belanja, analisis pertumbuhan pengeluaran membantu untuk memahami evolusi pengeluaran dari satu tahun ke tahun berikutnya. Terdapat penelitian sebelumnya yang menunjukkan nilai pertumbuhan pendapatan berpengaruh positif signifikan terhadap indikasi korupsi (Setyawan,2021);

c. Kerugian Keuangan Negara

Definisi keuangan negara sesuai UU No 17 Tahun 2003 tentang Keuangan Negara adalah, **“semua hak dan kewajiban negara yang dapat dinilai dengan uang, serta segala sesuatu baik berupa uang maupun berupa barang yang dapat dijadikan milik negara berhubung dengan pelaksanaan hak dan kewajiban tersebut”**. Selanjutnya, definisi kerugian keuangan negara berdasarkan Pasal 1 angka 22 UU No 1 Tahun 2004 tentang Perbendaharaan Negara adalah, **“kerugian negara/daerah adalah kekurangan uang, surat berharga, dan barang, yang nyata dan pasti jumlahnya sebagai akibat perbuatan melawan hukum baik sengaja ataupun lalai”**.

Berdasarkan uraian ketentuan perundang-undangan diatas, maka jelas bahwa yang dimaksud temuan kerugian keuangan negara dalam penelitian ini, adalah temuan auditor yang mengungkap permasalahan ketidakpatuhan terhadap peraturan yang berlaku dan mengakibatkan kerugian negara, contoh konkrit-nya antara lain adanya kurang pungut PNPB, pembayaran pekerjaan penyedia tidak sesuai dengan kuantitas dan/atau kualitas pekerjaan, dan sebagainya.

3. Metodologi

3.1. Identifikasi masalah dan Studi literatur

Pada penelitian ini, fokus masalahnya adalah bagaimana merancang sebuah pemrograman *machine learning* yang mampu memprediksi UPT yang berpotensi terdapat temuan kerugian negara berdasarkan kinerja keuangan.

Untuk studi literatur yang berkaitan penelitian ini, salah satunya yang menjadi literatur utama dalam pembuatan *machine learning* ini adalah hasil penelitian sebelumnya oleh Peneliti yang menunjukkan adanya pengaruh kinerja keuangan terhadap indikasi korupsi pada Kementerian, dimana indikasi korupsi adalah temuan kerugian negara sedangkan ukuran kinerja keuangan dimaksud terdiri dari nilai Likuiditas, Efisiensi Anggaran, Pertumbuhan Pendapatan dan Pertumbuhan Belanja.

3.2. Pengumpulan Data

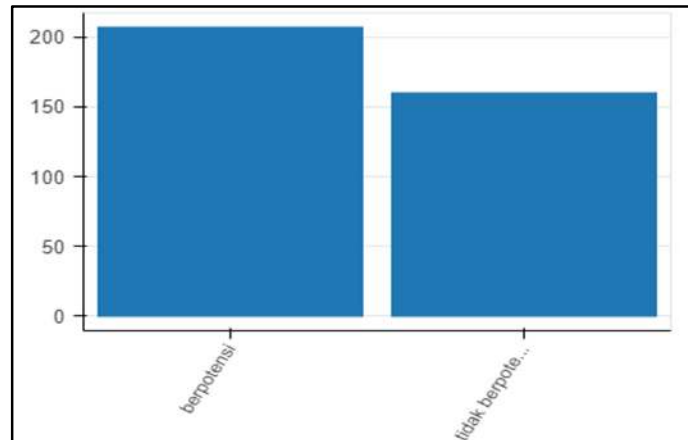
Pengumpulan data didapatkan dari dataset laporan keuangan dari situs resmi Direktorat Sistem Informasi Dan Teknologi Perbendaharaan kurun waktu 2016-2020 dan dataset Laporan Hasil Audit kurun waktu 2017-2021 dari Sistem Informasi Audit Inspektorat Jenderal Kementerian Perhubungan.

Informasi statistik data yang dikumpulkan:

Kolom	Record Non-Null / Non-duplicated	Type	KETERANGAN
Nama Auditee	367	object	Nama UPT
Propinsi	367	object	lokasi UPT
CR	367	object	rasio kas
ER	367	float64	rasio efisiensi
IIGR	367	float64	rasio pertumbuhan pendapatan

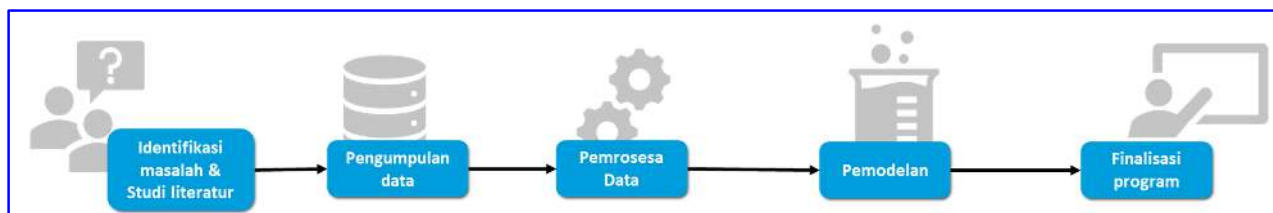
EGR	367	float64	rasio pertumbuhan belanja
Level Mod	367	object	Kolom informasi berpotensi atau tidak berpotensi temuan kerugian negara

Berdasarkan grafik sebaran data kolom informasi berpotensi atau tidak berpotensi temuan kerugian negara, diketahui dari 367 data yang dikumpulkan terdapat 56,4% adalah berpotensi dan 43,6% adalah tidak berpotensi



3.3. Perancangan

Tahapan perancangan digambarkan dalam diagram alir pada gambar.1 dibawah ini:



Gambar 1. bagan alir rancang bangun *machine learning*

3.4. Pemrosesan data

Pada tahapan ini, data-data yang telah dikumpulkan yaitu laporan keuangan (LK) dan laporan hasil audit (LHA) disandingkan sesuai dengan nama UPT-nya, selanjutnya dibuatkan *feature* (kolom) baru berdasarkan perhitungan kinerja keuangan yang terdiri dari:

a	Likuiditas	=	(kas + setara kas) kewajiban jangka pendek
b	Efisiensi Anggaran	=	$\frac{\text{realisasi pengeluaran}}{\text{realisasi penerimaan}}$
c	Pertumbuhan pendapatan	=	$\frac{(\text{pendapatan tahun berjalan} - \text{tahun sebelumnya})}{\text{pendapatan tahun sebelumnya}}$
d	Pertumbuhan belanja	=	$\frac{(\text{pengeluaran tahun berjalan} - \text{tahun sebelumnya})}{\text{pengeluaran tahun sebelumnya}}$

Pada tahapan ini, juga sekaligus memastikan hasil pemrosesan data tidak ada duplikasi dan nilai yang *error*

3.5. Pemodelan

Pada tahapan ini, dilakukan rancang bangun dan pengujian dari setiap pemodelan *machine learning*. Tipe pemograman *machine learning* untuk rancang bangun ini adalah algoritma

klasifikasi dan dibangun dengan bahasa pemrograman Python. Berdasarkan studi literatur sebelumnya terkait algoritma terbaik untuk pendeteksian fraud, maka diputuskan algoritma sebagai dasar pemodelan adalah *logistic regression*, *random forest* dan *k-Nearest Neighbors (KNN)*.

3.5.1 Cara kerja algoritma

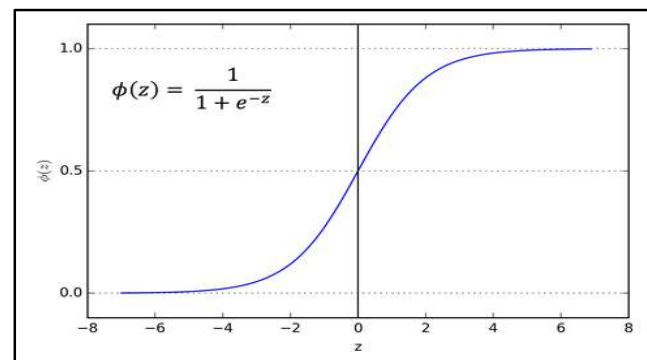
a. Logistic Regression

Logistic Regression adalah salah satu algoritma *machine learning* yang digunakan untuk masalah klasifikasi, khususnya untuk klasifikasi biner (dua kelas) di mana output yang diinginkan adalah nilai diskrit, misalnya "ya" atau "tidak", "benar" atau "salah", "positif" atau "negatif", dan sebagainya. Hasil dari *Logistic Regression* adalah probabilitas bahwa suatu data masukan termasuk dalam kelas yang ditentukan.

Cara kerja *Logistic Regression* adalah dengan menggunakan fungsi logistik (sigmoid) untuk memetakan input ke dalam rentang [0, 1], yang merepresentasikan probabilitas. Fungsi sigmoid digunakan untuk mengubah nilai-nilai linier dari fitur-fitur menjadi probabilitas yang berada dalam rentang tersebut. Rumus umum untuk *Logistic Regression* adalah:

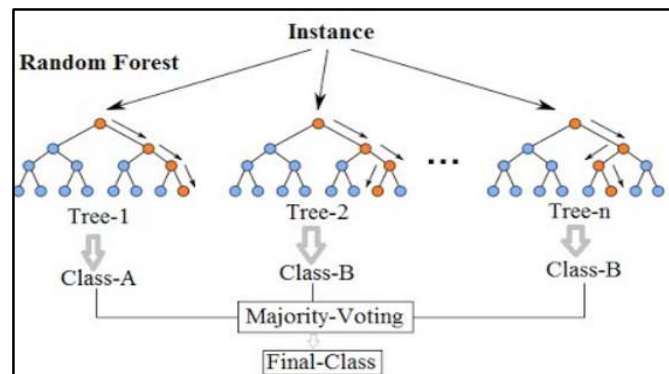
Di mana:

- z adalah hasil dari kombinasi linear dari fitur dan bobotnya.
- σ adalah fungsi sigmoid (fungsi logistik) yang mengubah z menjadi nilai probabilitas antara 0 dan 1.
- w adalah vektor bobot yang harus dipelajari selama proses pelatihan.
- x adalah vektor fitur input.



b. Random Forest

Random Forest bekerja dengan cara menggabungkan banyak pohon keputusan (*decision trees*) yang dibangun secara acak dari sampel data pelatihan, serta pemilihan subset fitur acak dari set fitur yang ada untuk setiap pohon keputusan. Kemudian, setiap pohon keputusan akan memberikan prediksi, dan hasil akhirnya diambil berdasarkan mayoritas suara (untuk klasifikasi) dari prediksi-prediksi tersebut.



c. k-Nearest Neighbors (KNN)

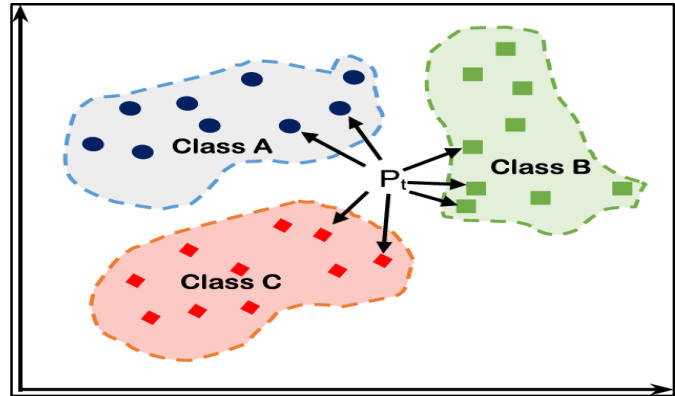
KNN bekerja berdasarkan prinsip bahwa data yang mirip cenderung berada dalam kategori yang sama atau memiliki nilai yang mirip.

Pada saat melakukan prediksi, KNN mencari *k-neighbors* (tetangga terdekat) dari data yang ingin diprediksi di antara data pelatihan yang sudah ada. "k" merupakan suatu nilai yang telah ditentukan sebelumnya dan merupakan jumlah tetangga terdekat yang akan digunakan untuk melakukan prediksi. Misalnya, jika $k=3$, maka algoritma akan

mencari 3 tetangga terdekat dari data yang ingin diprediksi, dan hasil prediksinya akan berdasarkan mayoritas label kategori dari ketiga tetangga tersebut (untuk klasifikasi) atau rata-rata nilai target dari ketiga tetangga tersebut (untuk regresi).

Salah satu kelebihan KNN adalah kemudahannya dalam implementasi dan dapat beradaptasi dengan data yang tidak terstruktur dengan baik.

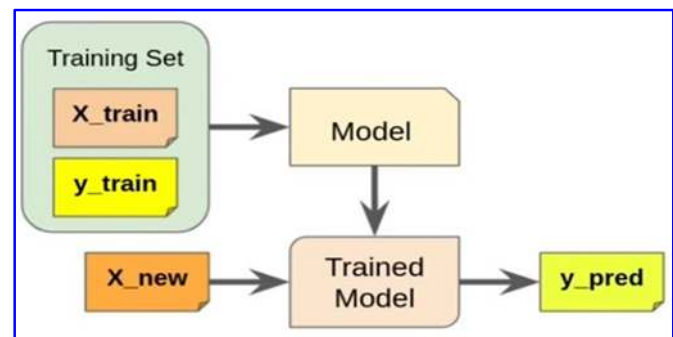
Namun, kelemahan utamanya adalah kinerjanya yang lambat ketika menghadapi dataset yang sangat besar karena harus melakukan perhitungan jarak untuk setiap titik data yang ada.



3.5.2 Hasil uji coba algoritma

Pengujian setiap algoritma pemodelan mengikuti alur diagram dibawah ini. Data yang telah dikumpulkan sebelumnya dibagi menjadi 70% untuk *training set*, sebagai bahan latih pemodelan. Sedangkan 30% data yang dikumpulkan sebagai bahan ujicoba untuk mengukur hasil prediksi. Hasil prediksi akan dituangkan dalam skor akurasi model.

Gambaran perhitungan akurasi model dapat dilihat pada gambar



	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} = \frac{\text{TN} + \text{TP}}{N + P}$$

dibawah ini:

Hasil pengukura akurasi ketiga model dengan setiap algoritma, sebagai berikut:

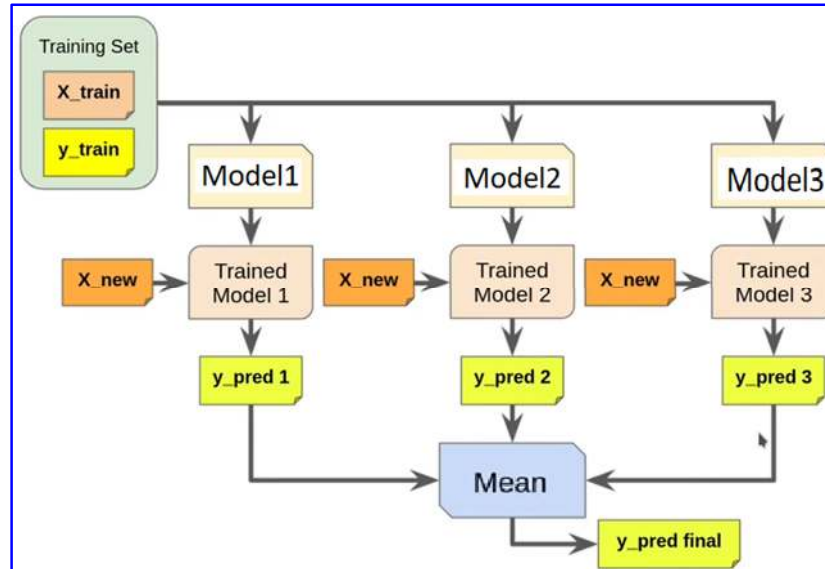
Model	Algoritma	Skor akurasi (%)
model1	<i>Logistic Regression</i>	73
model2	<i>Random Forest</i>	74

model3	<i>K-Nearest Neighbor</i>	65
--------	---------------------------	----

3.6. Finalisasi pemograman

3.6.1 Pemilihan model terbaik

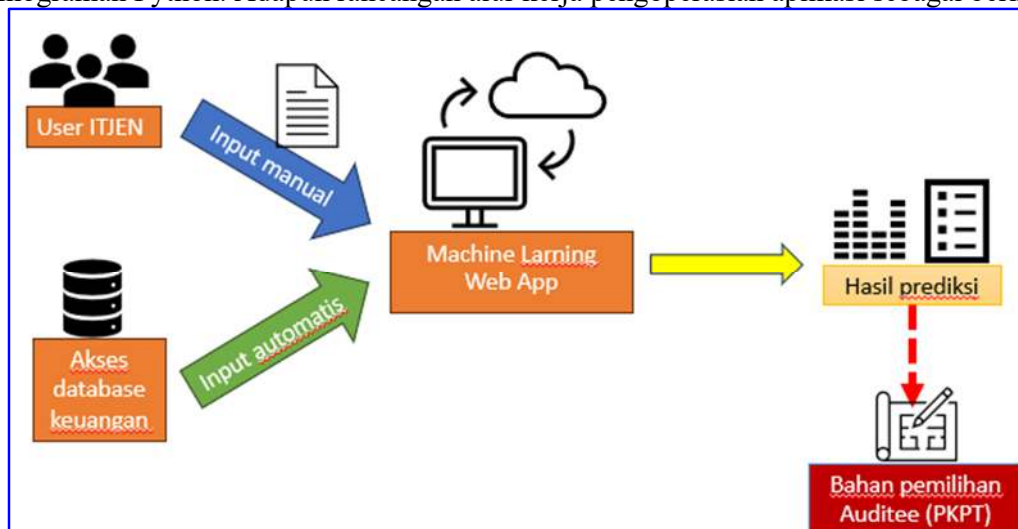
Berdasarkan studi literatur sebelumnya, bahwa ketiga algoritma yang dipilih merupakan algoritma terbaik dalam pendeteksian *fraud* sehingga untuk dapat menggunakan ketiga algoritma sekaligus untuk mendapatkan hasil prediksi yang paling optimal maka Peneliti menggunakan teknik *ensemble learning*, dimana setiap hasil dari ketiga model akan diambil nilai *mean* sebagai nilai akhir. Alur kerja penggabungan model sebagaimana gambar dibawah ini:



Gambar 2. bagan alir penggabungan ketiga model

3.6.1 Deployment Machine learning

Deployment dalam konteks *machine learning* mengacu pada proses saji agar *machine learning* dapat digunakan oleh publik. Pada tahap ini, Peneliti akan mengintegrasikan perancangan *machine learning* kedalam aplikasi berbasis website dengan bahasa pemrograman Python. Adapun rancangan alur kerja pengoperasian aplikasi sebagai berikut:



Gambar 3. Rancangan alur penggunaan dan pemanfaatan aplikasi website *Machine Learning*
 Prototipe aplikasi dapat dilihat pada link berikut: <https://auditee-prediction.streamlit.app/>

4. Hasil Akhir

Pada tahapan uji coba terakhir ini, Peneliti memasukkan data keuangan periode desember 2021 untuk memprediksi UPT yang berpotensi terdapat temuan kerugian keuangan negara yang akan muncul saat audit kinerja di periode tahun 2022, selanjutnya daftar UPT hasil prediksi tersebut akan dibandingkan dengan data aktual yaitu daftar UPT yang memiliki temuan kerugian keuangan negara berdasarkan laporan hasil Audit Kinerja periode TA 2022 yang telah rilis di Sistem Informasi Audit Inspektorat Jenderal Kementerian Perhubungan. Hasil perbandingan hasil prediksi dengan data aktual tersebut tersaji dalam tabel *confussion matrix* dibawah ini:

Jumlah data aktual 129		Data Aktual	
		BENAR ada temuan	TIDAK ada temuan
Hasil Prediksi	BENAR BERPOTENSI	52	12
	TIDAK BERPOTENSI	12	53

Keterangan tabel:

- Pada tahap uji akhir menunjukkan 105 UPT yang diprediksi tepat dari 129 UPT, sehingga didapatkan nilai akurasi prediksi sebesar $= (52+53)/129 = 81\%$.
- Dan diketahui 24 UPT salah prediksi yang terdiri dari:
 - 12 UPT sesuai hasil audit sebenarnya terdapat temuan, namun diprediksi tidak berpotensi;
 - 12 UPT sesuai hasil audit sebenarnya tidak terdapat temuan, namun diprediksi berpotensi

5. Penutup

5.1 Simpulan

Berdasarkan uraian sebelumnya dan hasil akhir, maka dapat disimpulkan sebagai berikut:

- Rancangan *machine learning* prediksi UPT yang berpotensi temuan kerugian keuangan negara berdasarkan kinerja keuangan dapat menjadi pertimbangan pemilihan Auditee sebagai bahan masukan PKPT Inspektorat Jenderal Kementerian Perhubungan;
- Kemampuan bagaimana *Machine Learning* dapat memprediksi sangat bergantung pada kualitas dan kuantitas data keuangan dan data hasil audit sebagai bahan latih pemodelan;
- Sebab kesalahan prediksi oleh *machine learning* dapat meliputi beberapa hal:
 - 1) variabel independen yang ada masih belum cukup
 - 2) Data yang dikumpulkan sebagai bahan latih pemodelan masih belum cukup,
 - 3) Temuan kerugian keuangan negara tidak signifikan untuk diungkap oleh Auditor
 - 4) Auditor kesulitan dalam mengungkap temuan kerugian keuangan negara pada UPT yang berpotensi.

5.2 Saran

- Sumber data untuk perancangan *machine learning* ini berasal dari Kementerian Perhubungan, sehingga sangat disarankan jika hendak diimplementasikan diluar Kementerian Perhubungan maka sumber data sebagai bahan latih pemodelan menyesuaikan sasarannya;

- b. Masih perlu melakukan penelitian lanjutan untuk mencari variabel independen baru yang dapat membantu meningkatkan akurasi prediksi *machine learning* ini.

6. Daftar Pustaka

- Setyawan, Eka C.(2021). The Effect Of Financial Performance On Corruption Indications With Internal Control System As Moderation At The Ministry Of Transportations
- Brownlee, J. (2020). Machine Learning Mastery with Python. eBook.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer Science & Business Media.
- Alpaydin, E. (2010). Introduction to Machine Learning. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
- Peraturan Pemerintah Republik Indonesia Nomor 60 Tahun 2008 Tentang SPIP
- Durtschi, C., Hillison, W., dan Pacini, C., (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. Journal of Forensic Accounting.
- Tuanakotta, M.Theodorus. 2007. Akuntansi Forensik dan Audit Investigatif, edisi pertama,
- Undang-Undang Republik Indonesia Nomor 17 Tahun 2003
- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- Mitchell, T. M. (1997). Machine Learning. McGraw Hill.
- "Deployment of Machine Learning Models" di Towards Data Science
<https://towardsdatascience.com/deployment-of-machine-learning-models-part-1-2acf3cf9b9b2>
- "Machine Learning Deployment Tutorial" di Machine Learning Mastery
<https://machinelearningmastery.com/deploy-machine-learning-model-to-production/>