

NEW YORK STOCK EXCHANGE

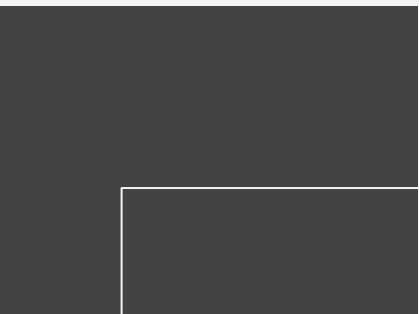


ALONA

ALokasi Outcome aNomAli

Presented by GASPOL Tech





Belum ada pengukuran outcome APBD yang resmi digunakan dalam menentukan alokasi anggaran berdasarkan fungsinya, pengukuran ini penting untuk melihat apakah belanja sudah efektif dan efisien.



Latar Belakang



ALokasi
Outcome
aNomAli



ALONA

Alona mencoba untuk melakukan **pengukuran outcome belanja APBD** yang berdasarkan pada **Index Pembangunan Manusia** (metode baru) yang dikeluarkan oleh BPS menggunakan teknik **Machine Learning** dan melakukan *deployment* dalam bentuk **website** menggunakan *opensources tools*.



Tujuan dan Manfaat ALONA

Memahami keterkaitan belanja APBD berdasarkan fungsi dengan IPM

01

Membuat model prediksi IPM dengan metode Machine Learning

02

Memilih algoritma optimal untuk prediksi indikator IPM untuk deployment

03

Membuat dashboard untuk realokasi anggaran berdasar outcome

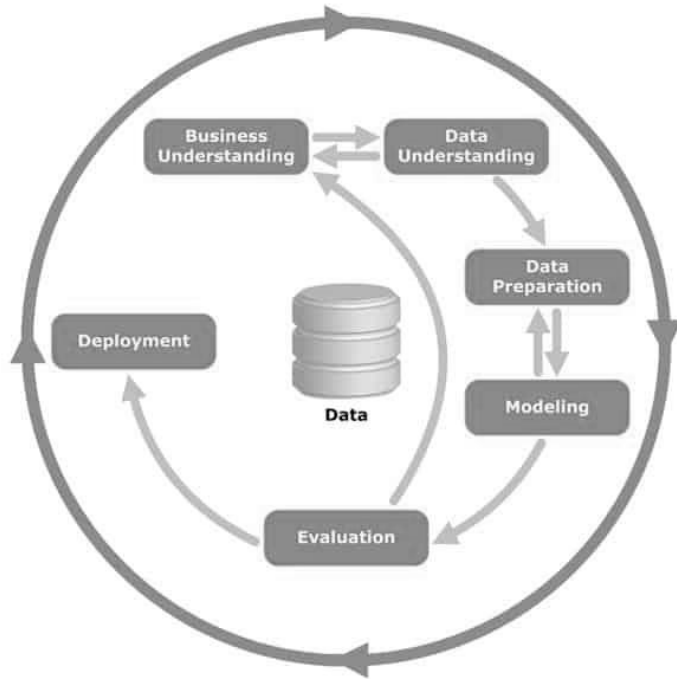
04

Menambah fitur dashboard milik DJPK untuk mencari outcome optimal

05



Metodologi



CRISP-DM

The Cross Industry Standard Process for Data Mining **paling tepat** digunakan untuk project ALONA ini, karena diperlukan pengulangan dalam persiapan data dan modelling untuk mendapatkan **akurasi prediksi yang terbaik** yang bisa didapatkan dari dataset yang disediakan.

IPM is the ideal predictor so far

Dimensi	Metode Lama		Metode Baru	
	UNDP	BPS	UNDP	BPS
Kesehatan	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)
Pengetahuan	Angka Melek Huruf (AMH)	Angka Melek Huruf (AMH)	Harapan Lama Sekolah (HLS)	Harapan Lama Sekolah (HLS)
	Kombinasi Angka Partisipasi Kasar (APK)	Rata-rata Lama Sekolah (RLS)	Rata-rata Lama Sekolah (RLS)	Rata-rata Lama Sekolah (RLS)
Standar Hidup Layak	PDB per kapita (PPP US\$)	Pengeluaran per kapita Disesuaikan (Rp)	PNB per kapita (PPP US\$)	Pengeluaran per kapita Disesuaikan (Rp)
Agregasi	Rata-rata Aritmatik $IPM = \frac{1}{3}(I_{\text{kesehatan}} + I_{\text{pendidikan}} + I_{\text{pengeluaran}}) \times 100$		Rata-rata Geometrik $IPM = \sqrt[3]{I_{\text{kesehatan}} \times I_{\text{pendidikan}} \times I_{\text{pengeluaran}}} \times 100$	

Data Preparation is the biggest challenge

- Cleansing 6 juta dataset mengelompokkan ke fungsi belanja, kemudian melakukan agregasi per daerah
- Pengambilan data IPM dan jumlah penduduk melalui situs BPJS provinsi dilakukan secara manual.
- Banyak data yang tidak lengkap di data BPJS.
- Perbedaan nomenklatur penamaan daerah antara DJPK dengan BPS



Modelling

Modelling dilakukan menggunakan Python, dengan data engineering sederhana, yaitu membagi jumlah belanja dengan jumlah penduduk (per kapita), dimana meningkatkan akurasi prediksi. Model yang digunakan ada 13 macam model mulai dari Ordinary Least Square hingga Stack atau Ensemble, dengan target IPM dan komponennya.

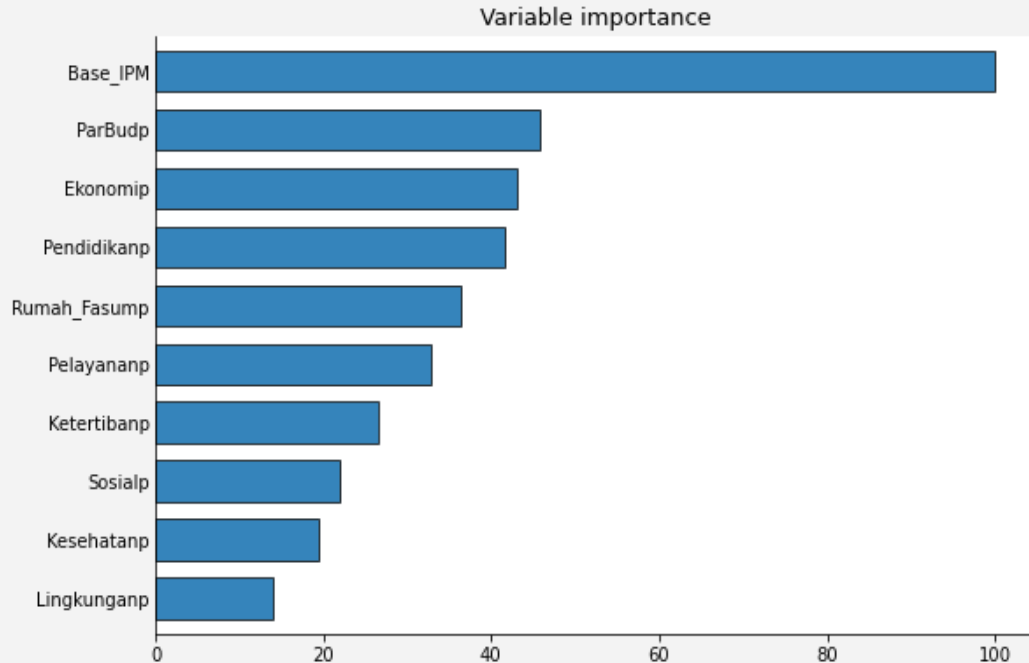
Hasil dan Evaluasi

	Test RMSE	Test R2	Test MAE
OLS	0.171	0.162	0.144
KNN	0.190	-0.033	0.159
Lasso	0.183	0.034	0.154
Ridge	0.189	-0.031	0.155
Elastic Net	0.183	0.034	0.154
Tree	0.240	-0.657	0.195
Bagged Trees	0.184	0.024	0.153
Random Forest	0.177	0.099	0.148
Boosting	0.177	0.097	0.145
LightGBM	0.189	-0.029	0.155
XGBoost	0.188	-0.014	0.152
Additive Boost	0.187	-0.007	0.149
Stack	0.177	0.102	0.147

Hasil dari setiap indikator berbeda model untuk setiap akurasi terbaiknya, ALONA yang dapat dilihat pada di kiri (IPM) dan keseluruhan (table bawah). Meskipun begitu kami memilih model yang paling cocok feasible dilakukan deployment pada streamlit. Contoh untuk IPM kami menggunakan Gradient Boosting disbanding OLS, karena tidak terkendala saat deployment.

Indikator	Rerata Asolute	MAE	MAE : MeanA	Model
AHH	0.28704611	0.119	41%	Stack
HLS	0.083717579	0.082	98%	KNN
RLS	1.271815562	0.091	7%	Ridge
PPK	355.2175793	107.607	30%	Stack
IPM	0.644610951	0.144	22%	OLS

Insight menarik dari hasil Machine Learning – IPM (Gradient Boosting)

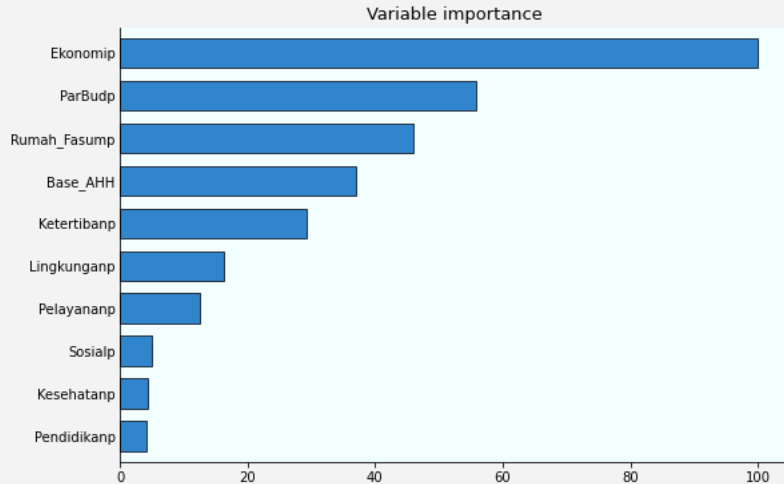


Meskipun variabel importance beragam untuk setiap model, kami mengambil elemen penting dari prediksi untuk masing-masing indicator. IPM selain dari base IPM, variabel terpenting justru ada pada Pariwisata dan Budaya, dibanding dengan Pendidikan maupun Ekonomi dan Kesehatan.

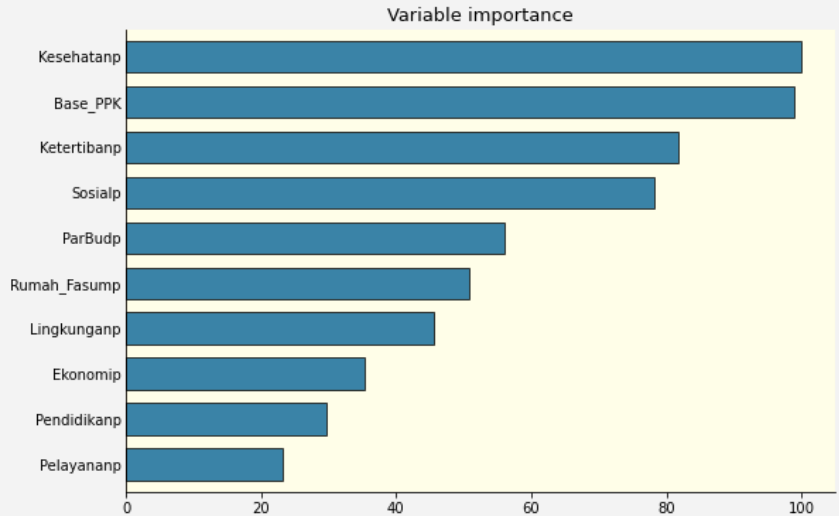
Hal ini menunjukkan bahwa belanja pada sektor selain fungsi Kesehatan, Pendidikan, dan Ekonomi memiliki korelasi yang tidak kalah besar dalam meningkatkan IPM.



Insight menarik dari hasil Machine Learning – AHH dan PPK (Gradient Boosting)



Pada Angka Harapan Hidup, belanja ekonomi ternyata memiliki variable importance yang lebih tinggi disbanding dengan Kesehatan yang ada di urutan dua terbawah.



Sedangkan pada Pendapatan per Kapita, factor belanja Kesehatan lebih penting disbanding dengan belanja ekonomi.

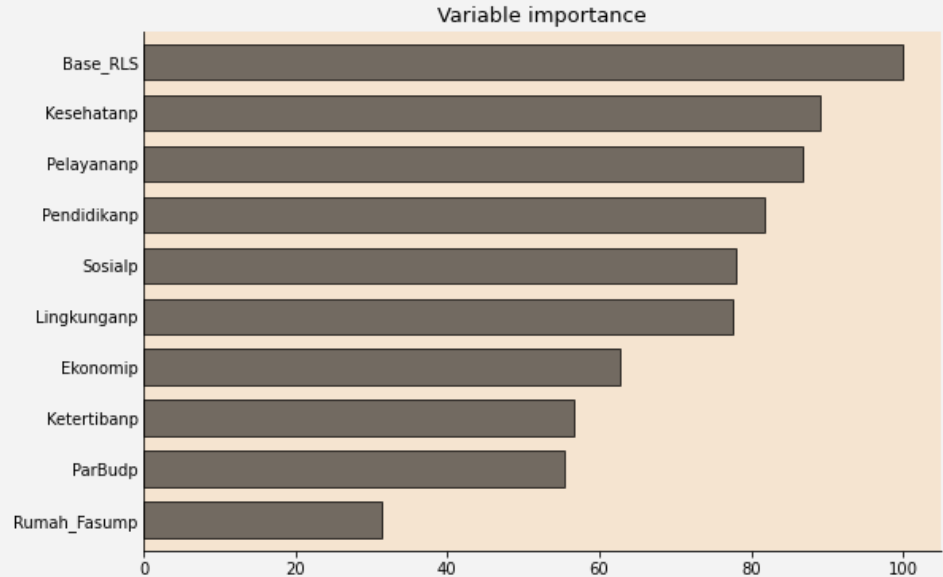


Insight menarik dari hasil Machine Learning – RLS (Gradient Boosting)

Sedangkan pada Rata-rata Lama Sekolah, selain dari nilai dasar RLS, belanja Kesehatan dan pelayanan memiliki variable importance yang lebih tinggi dibandingkan Kesehatan.

Ada dua kemungkinan hal ini terjadi, yang pertama data dan model tidak secara memadai dalam melakukan prediksi, dan yang kedua adalah outcome dari belanja tidak sesederhana melakukan fokus belanja pada fungsi yang terkait dengan indikator atau indeks.

Oleh karena itu, selain perlunya perbaikan model di masa depan, kami juga mencoba membuat dashboard yang dapat memprediksi outcome dari komposisi belanja berdasarkan fungsi.



ALONA (Login page: gaspoltech.com)

gaspoltech.com/login/?next=/

ALONA Bedah Data APBD

Add your credentials

Username

Password

Remember Me

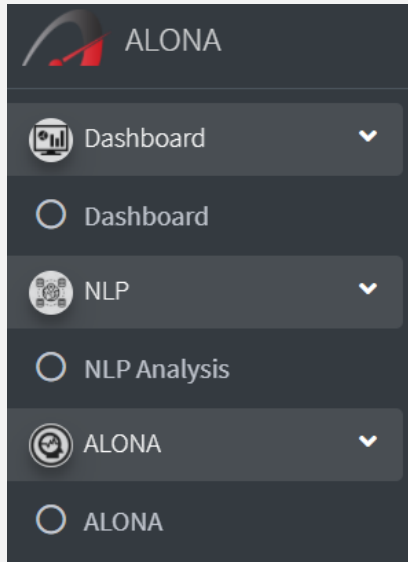
Not a member? [Register](#)

© Gaspol

- Alona dapat di-*deploy* pada hosting website local
- Alona memiliki fitur user management sehingga akan memudahkan DJPK apabila ingin menjadikan pemda-pemda sebagai user untuk melihat dashboard.
- Alona menggunakan open source Python, plotly dan streamlite, sehingga peningkatan scalability dapat dilakukan dengan lebih murah dibandingkan menggunakan software berbayar seperti Tableau.



Menu ALONA



Ada tiga menu utama dari ALONA selain dari core Alona itu sendiri, yaitu:

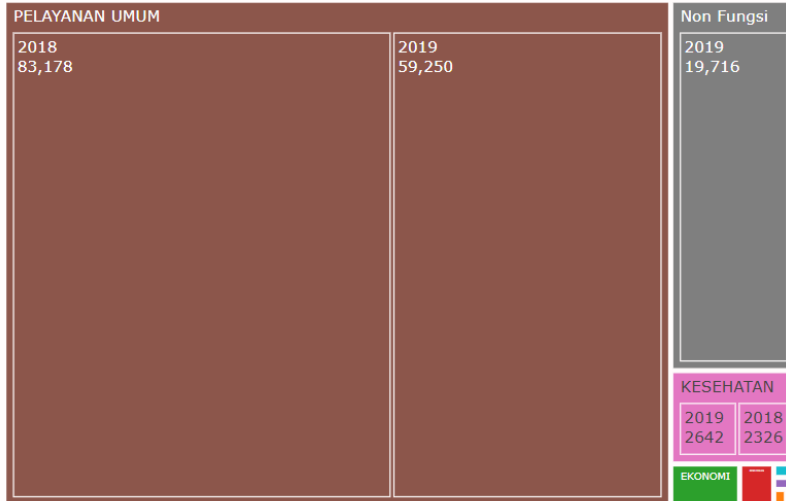
- **Dashboard:** visualisasi menggunakan 6,7 juta data raw yang diberikan oleh DJPK
- **NLP Analysis:** Membuat wordcloud dengan beberapa cluster untuk mengetahui semisal apa saja yang terkait dengan belanja rutin.
- **ALONA Core:** Melakukan prediksi dengan Machine Learning untuk melakukan pengukuran outcome atas nilai dan komposisi belanja daerah berdasar fungsi. Melakukan visualisasi untuk melihat apakah ada anomaly antara pengeluaran dengan peningkatan outcome di beberapa daerah.



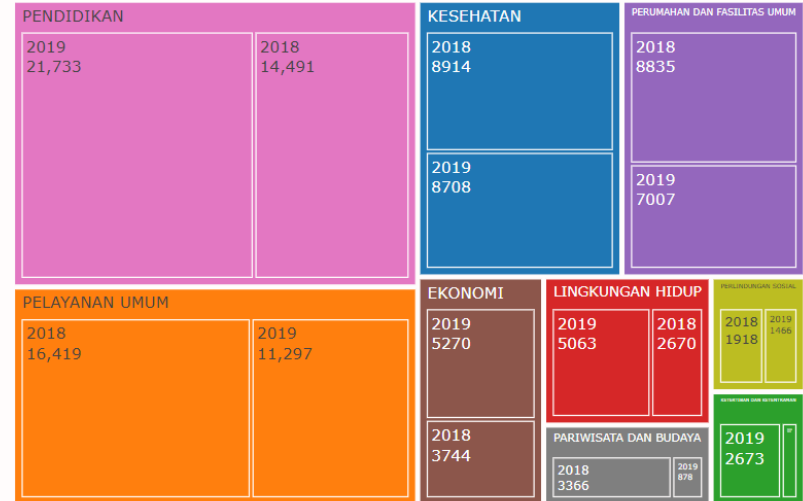
Dashboard Visualisasi – 6,7 jt rows



Penerimaan Per Fungsi (Dalam Milyar Rp)



Pengeluaran Per Fungsi (Dalam Milyar Rp)



<input type="checkbox"/>	Kalimantan Tengah	33138	20543714406581.883
<input type="checkbox"/>	Kalimantan Timur	44815	45584185864195.14





Prediksi Kluster Anggaran Dengan NLP

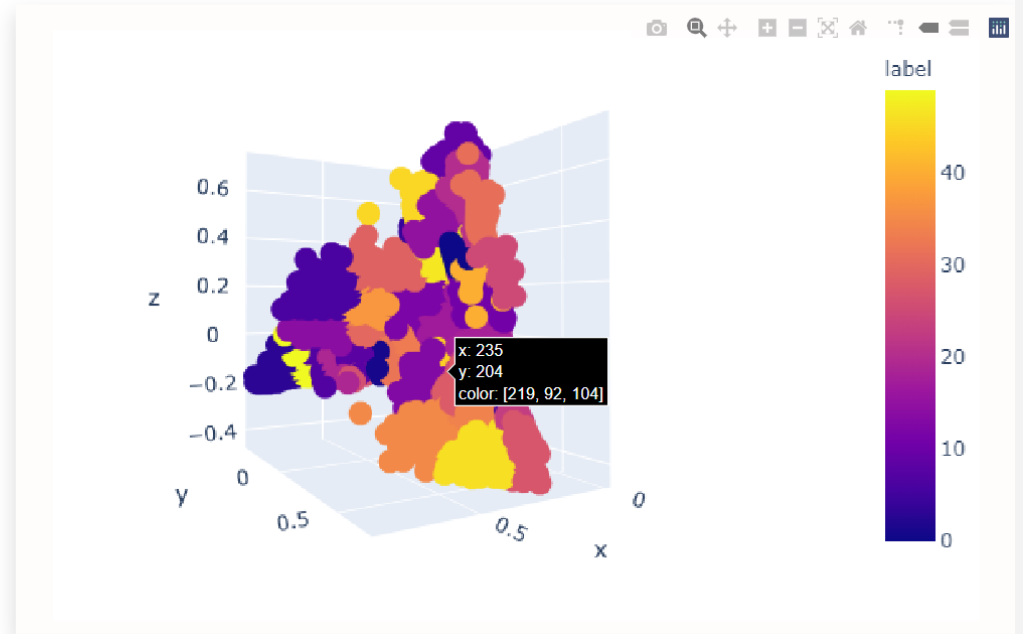
Masukkan Kata Kunci

Submit Text

Kluster Prediksi: None

Kluster Anggaran

Sebaran Kluster



Dashboard ALONA (Anomali)



Pilih Index

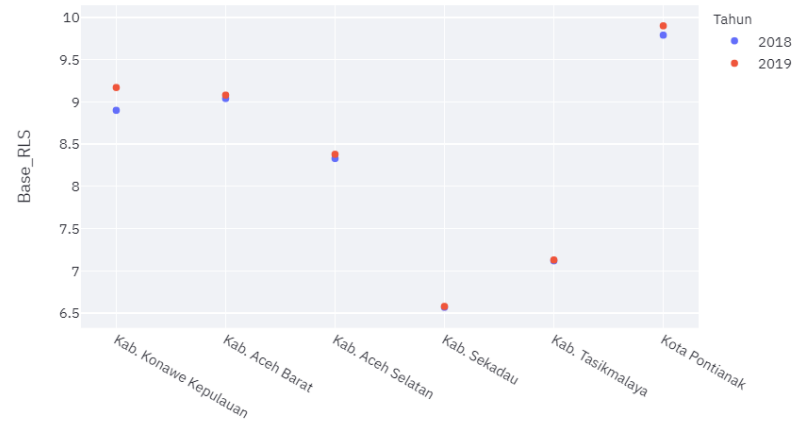
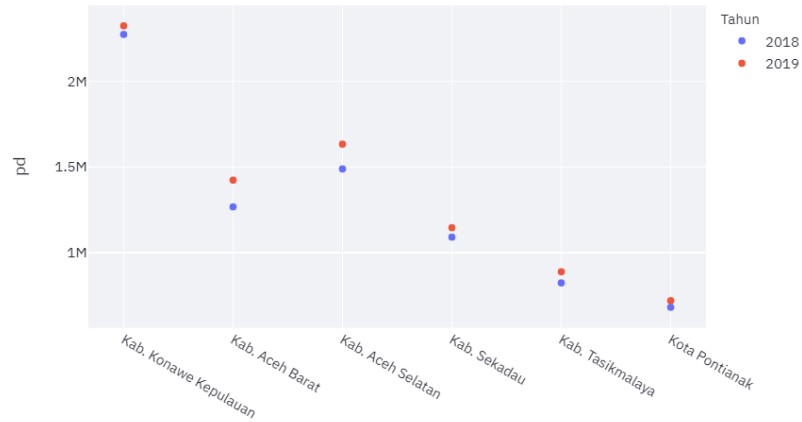
RLS

Pilih Jenis Fungsi

Pendidikan

Pilih Pemda

Kab. Aceh Barat ✕ Kab. Aceh Selatan ✕ Kab. Konawe Kepulau... ✕ Kab. Sekadau ✕ Kab. Tasikmalaya ✕ Kota Pontianak ✕



Dashboard ALONA (Comparison)



Select Type of Plot

line

Pilih Pemda

Kab. Aceh Barat Daya X

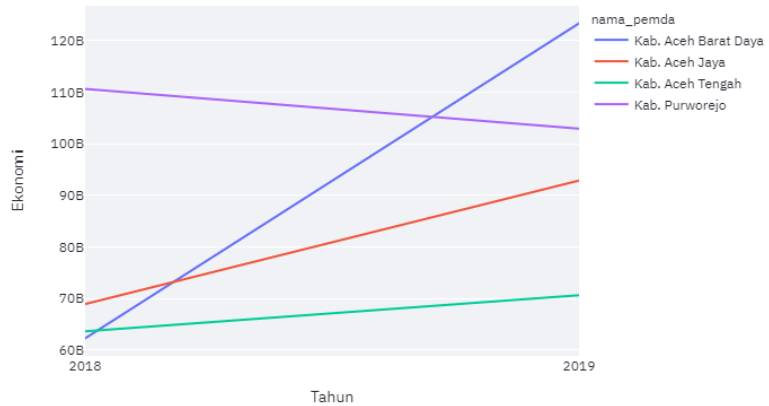
Kab. Aceh Jaya X

Kab. Aceh Tengah X

Kab. Purworejo X

Select Column To Visualize

Ekonomi



Dashboard ALONA (Alokasi dan Outcome)



Select Menu

Allocation & Outcome

ALONA: Allocation - Outcome - Anomali

Alokasi Anggaran dan Prediksi Perubahan Index

Pilih Jenis Wilayah

Kab_Kota

Pilih Daerah

Kab. Aceh Selatan

Total Populasi

238081

(Populasi Penduduk Tahun 2019)

Total Anggaran Tahun Terakhir

1206764324216

Anggaran Perkapita

5068713

% Penyesuaian Anggaran

10

Total Anggaran Akan Dialokasikan

1327440756637

Anggaran Perkapita Akan Dialokasikan: 5575584,60

Alokasi Tahun Sebelumnya (Rp)

Ekonomi

116382569647

Kesehatan

283011202514

Ketertiban

14525297950

Lingkungan

13021826876

Pariwisata Budaya

Proporsi Alokasi Anggaran (%)

Ekonomi

9.64

Kesehatan

23.45

Ketertiban

1.20

Lingkungan

1.08

Pariwisata Budaya

Jumlah Anggaran Dialokasikan (Rp)

Ekonomi

128020826611

Kesehatan

311312322765

Ketertiban

15977827744

Lingkungan

14324009563

Pariwisata_Budaya

Prediksi Perubahan Index

Berdasarkan Machine-Learning Models

Pilih Index

IPM

Nilai IPM Awal

Base_IPM

65.92

Total Anggaran Dialokasikan: 100.00 %

[Klik disini untuk mendapatkan Nilai Prediksi](#)

Prediksi Index: 66.72

Keterbatasan dan Kesimpulan Proposal ALONA



- Keterbatasan tenaga dan waktu pengerjaan serta dataset
- Model masih sederhana dan belum memiliki akurasi yang diharapkan
- Perlu dilakukan penelitian lebih lanjut terkait fungsi dan outcome APBD
- Diharapkan dapat membuka insight baru dalam pendekatan alokasi anggaran



Pengembangan Lanjutan

**Menambah dataset
minimal 5-10 tahun
terakhir**

**Bekerjasama dengan
BPS terkait indikator
outcome yang lebih
tepat**

**Perbaiki model misal
pengelompokan fungsi
dan jenis belanja
(modal-barang)**

**Perbaiki UX UI
dashboard ALONA agar
lebih mudah digunakan
user**



Ginanjar

Lead Programmer
and
Machine Learning
Expert



Agung

Visualization and
Database Expert



Sindhu

Analyst and UX UI
Expert

GASPOL Past Projects:

- **Fismart** (Fishery Informatics and Smart Marketplace) - **1st Winner** of Maritim Hackathon (2021)
- **PMD** (Preventive Measures Dashboard) - **1st Winner** of BPJS Visualthon (2020)
- **Hapekita** (Multi Verification Phone Sharing Apps) - **Finalist** of Japan Hackathon (2021)
- **IGUT** (Insight Generator for Unstructured Text) - **Finalist** of BPK Hackathon (2021)
- **GEDS** (Geographical Epidemiology Decision Support) - **Semi Finalist** of Microsoft Azure Hackathon (2021)



Thanks

Does anyone have any questions?