



Account Nomenclature Standardization and Expenditure Effectiveness Analysis of Regional Budget in Indonesia

**YOLO - Winners of 2022 Indonesia Better Budget
Dataquest “Bedah Data APBD 2022 DJPK”**

Sharing Session on March 7th, 2023 at
“**Learn from the champions! Join the international
showcases of fiscal data usage champions**”
Global Initiative for Fiscal Transparency



YOLO (You Only Live Once)



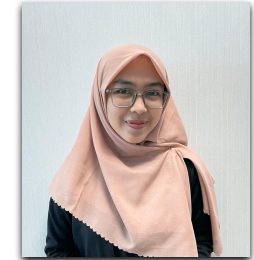
Aldo Lefiro Irfiansyah
<https://www.linkedin.com/in/leofirol>
aldo.lefiro@gmail.com



Muarrikh Yazka Ginang P
<https://www.linkedin.com/in/muarrikhyazka/>
muarrikhyazka@gmail.com



Nashir Wahyudi
www.linkedin.com/in/nashir-geeks-wahyudi
nashirwahyudi@gmail.com



Novia Permatasari
https://linktr.ee/viasari_
noviaprmatasari@gmail.com

Our team journey started in 2020 ...



2017



2020



2021



2022

*We believe this space
will be full of another
journey &
accomplishments*

Track Record (19 Competitions and much more to come . . .)

1. **1st Winner of *Bedah Data 2022* by Ministry of Finance, Indonesia**
Standardization of Account Nomenclature and Analysis of the Effectiveness of National Budgeting Expenditures in Indonesia
2. **3st Winner of *Bedah Data 2022* by Ministry of Finance, Indonesia**
Standardization of Budget Nomenclature and Prediction of Strategic Indicators Achievement
3. **3rd Winner of *Regsosek Data Festival 2022***
Dashboard Analysis of Family Welfare Profiling Using Machine Learning and Geospatial Analysis.
4. **2nd Winner of *Jaga SPI Data Challenge 2022* by Corruption Eradication Commission**
Proxy for the Ministry/Institutional Integrity Index using Big Data
5. **Finalist of *UN Big Data Hackathon 2022***
Prediction of Food Security Index Using Big Data
6. **2nd Winner of *Jaga Data Challenge 2021***
Mantau BOS, Anomaly Detection and Optimization of School Operational Assistance Fund in Indonesia
7. **3rd Winner of *Kemenkeu Data Hackathon ASN 2021***
Use Case Realisasi Anggaran
8. **Finalist of *Global AI Innovation Challenge Series 2021***
Intelligent Weather Forecast for Better Life
9. **Finalist of *BPJS Hackathon 2021* – Artificial Intelligence**
10. **2nd Winner of *Hackathon DJP IT Summit 2021* – VAT and Cross Border Fraud Detection**
11. **Finalist of *Hackathon DJP IT Summit 2021* – Tax Court Verdict Prediction**
12. **4th Winner of *Bedah Data APBD 2021* at DJPK Kemenkeu**
13. **Finalist of *Data National Hackathon* at Data Science Weekend 2021**
14. **2nd Winner of *BRI Data Hackathon 2021*, People Analytics**
15. **Finalist of *ASEAN BPK Hackathon 2021***
16. **Finalist of *HackData Indosat Ooredoo 2020***
17. **Finalist of *Jenius Hackathon 2020***
18. **2nd Winner of *Datathon 2019* at LKPP, KPK, ICW**
19. **4th Winner of *HackData Indosat Ooredoo 2019***



Why we interested in using budget data

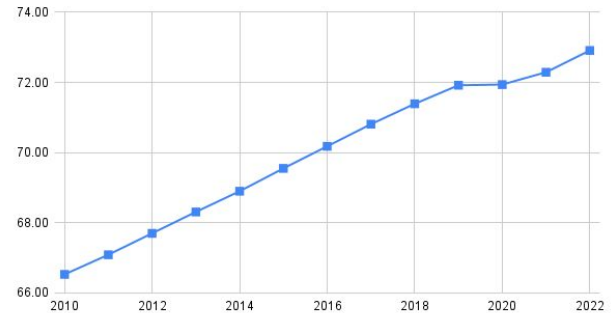
Basic Reason

- Purely wondering about this data
- Love experimenting and challenge new problems using data science
- Contribute to society :D

Situational Reason

- Slow growth on HDI 2019 onwards, so it needs more attention. We believe that **proper budget plan will have a good impact on the region's development.**

Human Development Index (2010-2022)



What's our project?

Project Goals

- **Standardize the number level 6 account nomenclature**, based on the Minister of Home Affairs 050-5889 in 2021.
- Provide **recommendations for the allocation of Regional Revenue and Expenditure Budget (APBD)** based on the Machine Learning model.



Methodology

Standardization of level 6 Account Nomenclature



Government Budgeting Optimization

Text Processing

Pre-trained Indonesian BERT

Semantic Search

Cosine Similarity

Clustering & Profiling Analysis

K-Means Clustering

HDI Prediction & Optimum Allocation

Regres & SHAP Values

Project's Output - Standardization

Example of Semantic Search Result

Keyword :

“Belanja Iuran Jaminan Asuransi”

Result with the highest similarity based our standict model is ***“Belanja Iuran Jaminan Kesehatan PNS”*** with account code **‘5.1.01.01.09.0001’** and similarity score **0.886**.

Kata Kunci : Belanja Iuran Jaminan Asuransi

Hasil Semantic Search:

Batches: 100%  1/1 [00:00<00:00, 9.14it/s]

	Nama Akun	Cosine Similiarity Score	kode_akun
0	belanja iuran jaminan kesehatan pns	0.885833	5.1.01.01.09.0001
1	belanja iuran jaminan kematian pns	0.877256	5.1.01.01.11.0001
2	belanja iuran jaminan kecelakaan kerja pns	0.853252	5.1.01.01.10.0001
3	belanja iuran jaminan kematian pppk	0.836419	5.1.01.01.11.0002
4	belanja iuran jaminan kesehatan bagi non asn	0.828114	5.1.02.02.02.0005 b
5	belanja iuran jaminan kesehatan pppk	0.823282	5.1.01.01.09.0002
6	belanja iuran jaminan kecelakaan kerja pppk	0.818283	5.1.01.01.10.0002
7	belanja iuran jaminan kesehatan bagi dprd	0.813440	5.1.01.04.12.0001
8	belanja iuran jaminan kesehatan bagi mrp	0.809018	5.1.01.07.12.0001
9	belanja sewa takaran lainnya	0.807665	5.1.02.02.04.0095

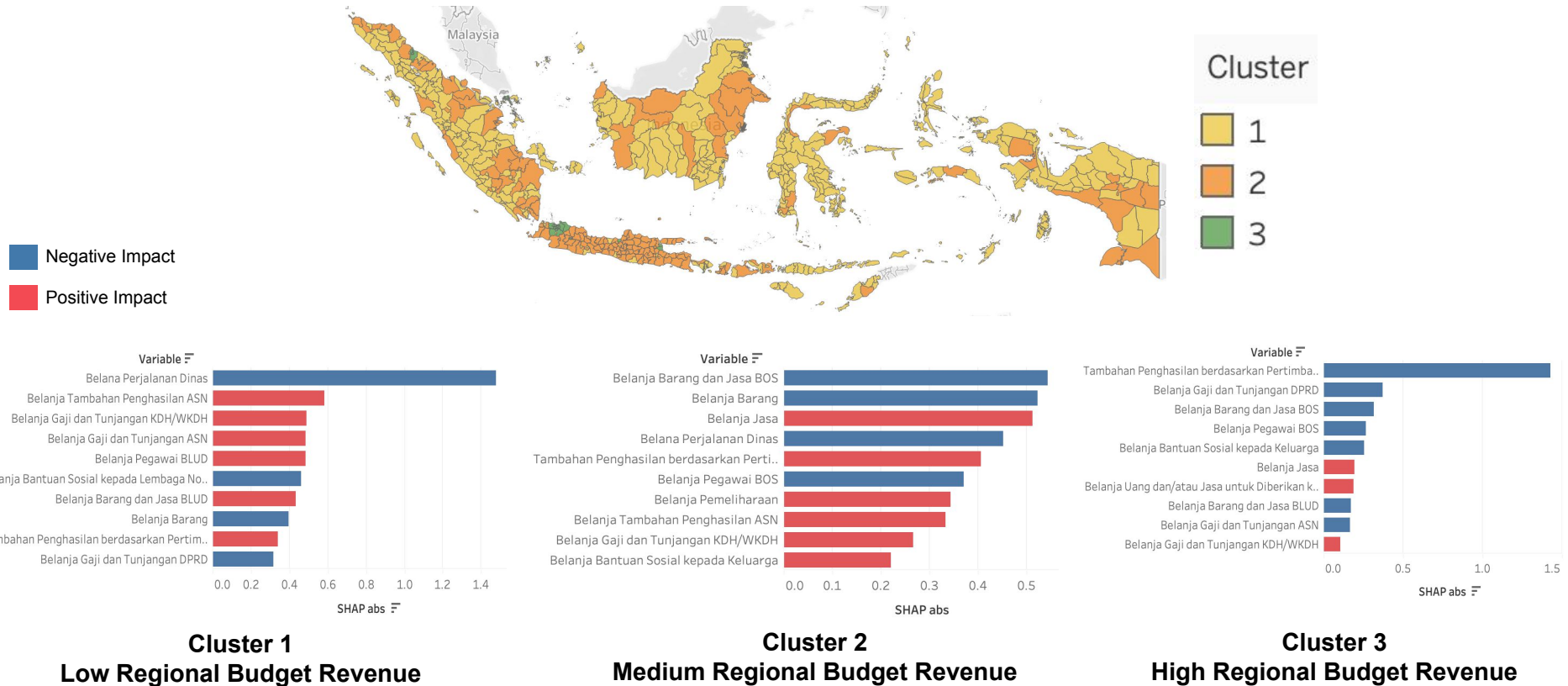
Demo



standict
apps



Project's Output - Budgeting Optimization



Recommendation for Policy Maker

- System standardization and integration uses Standict on regional government so that **account code nomenclature errors can be minimized or even eliminated.**
- **Allocating the APBD operational expenditure budget based on the account characteristics of the regional income cluster** formed by the optimizing the use of the APBD on accounts that have a positive or productive influence on contributions to HDI (Human Development Index) in an area.

Suggestion

- Implement **one standard** for account code and other code in data for better data analysis.
- Open up **the access** to budget data into public so everyone can join analyzing and putting idea about it.
- Push technology in every stage of budgeting, so it will be more **effective** and **efficient**.
- Create more competitive modeling like kaggle competition :)

Thank You!

#DataMencerdaskanBangsa
#MencatatPertanianIndonesia
#MencatatUntukMembangunNegeri

#MenjagaKeuanganNegara

Standardization of level 6 Account Nomenclature

SEMANTIC SEARCH

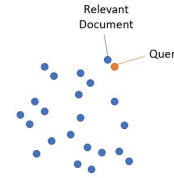
"Semantic search seeks to improve search accuracy by **understanding the content** of the search query. In contrast to traditional search engines, which only find documents based on lexical matches, semantic search can also find synonyms."

BERT

- **BERT** is a method/mechanism that aims to **convert input text into numerics** so that modeling can be done.
- **BERT - Base Indonesia** is an Indonesian version of the Bert model that is trained using words derived from Indonesian Wikipedia in the form of Masking Language Model.
- **Tuning** is done to make the pre-trained model in accordance with the data dictionary to be used, budgeting data.

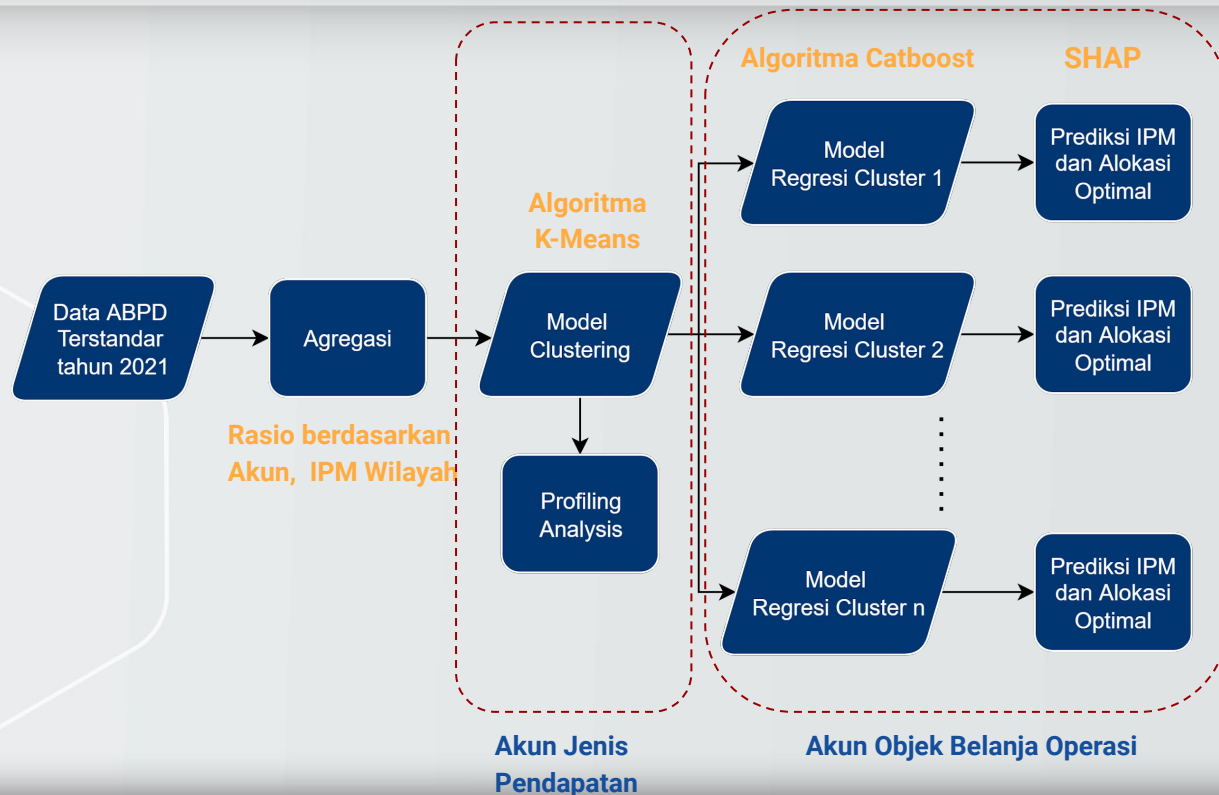
Cosine Similarity

- The same or nearest field should have a high semantic/meaningful similarity with the queries sought.
- The values of the closeness of the two queries are calculated using the **Cosine Similarity** formula:



$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Expenditure Effectiveness Analysis of Regional Budget



We used the **K-Means method** based on standardized data for grouping the regions.

Regression modeling using catboost on each cluster to see the effect of allocation. At this stage additional data is used, namely the **Human Development Index** as a basis for seeing the impact of the budget allocation applied to certain regions.

Sistem Usulan

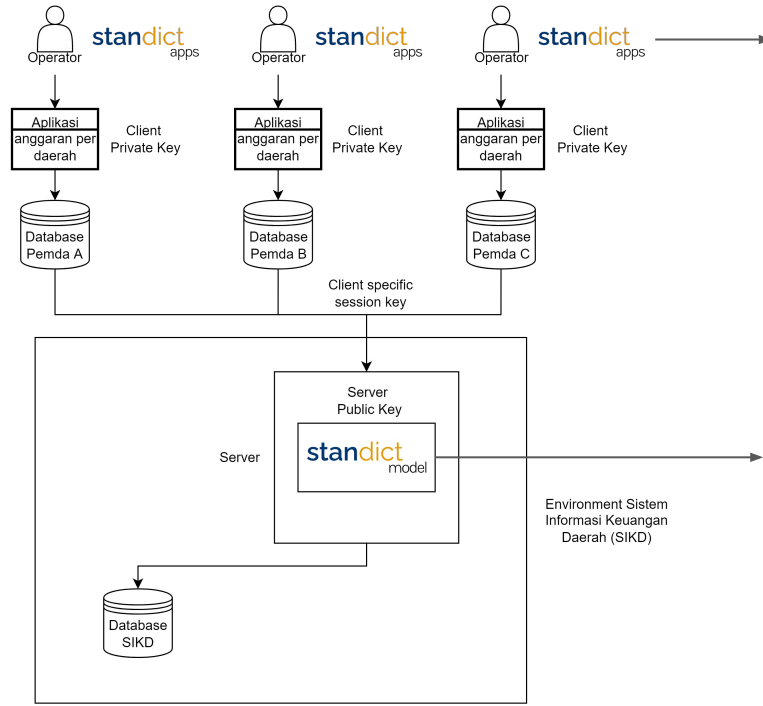
standict
apps

Standardization Dictionary (standict) Apps adalah **aplikasi** berbasis web yang menjadi kamus dengan kapabilitas pencarian berbasis **semantic/arti kata** dari nomenklatur subkegiatan

standict
model

Standardization Dictionary (standict) Model adalah **model** yang telah dilengkapi dengan kapabilitas pencarian berbasis **semantic/arti kata** dari nomenklatur subkegiatan

Arsitektur Sistem Usulan



Standict Apps membantu Operator atau penyusun APBD di daerah dalam **menentukan** nomenklatur sub kegiatan yang baku

Standict Model membantu DJPK untuk **menvalidasi dan memperbaiki** subkegiatan dari data APBD yang telah disubmit oleh daerah agar sesuai dengan standar sebelum masuk ke Database SIKD