

ANALITIKA DATA UANG PERSEDIAAN SATUAN KERJA PENGGUNA APBN

achmad ford¹, siti fatimatus zahroh², rachel dwi aurindah³

¹²³ Tim Data101 Kantor Pelayanan dan Perbendaharaan Negara Sumbawa Besar

Abstrak

Uang Persediaan adalah uang muka kerja yang dikelola oleh Satker dalam rangka pelaksanaan anggaran. Dalam rangka bahan rujukan bagi berbagai pihak yang terlibat dalam pengelolaan Uang Persediaan sesuai dengan peran pihak yang berkepentingan, akan dilakukan sebuah penelitian dengan menggunakan teknik analisis deskriptif kuantitatif yang didukung oleh visualisasi data. Metode penelitian ini menggunakan pendekatan Design Science Research Methodology (DSRM). Proses analitika akan mencari insight data yang akan bermanfaat untuk pemodelan analitika data baik deskriptif dan prediktif serta melakukan penalaran hasil dari insight dan model yang dihasilkan agar dapat dimanfaatkan kedepannya. Dari hasil analitika deskriptif diperoleh informasi Kanwil dan KPPN yang mengelola Uang Persediaan dengan baik yang dapat dijadikan tolak ukur pengelolaan Uang Persediaan oleh Kanwil dan KPPN lain. Analitika prediktif menunjukkan pengelompokan satker pengelola Uang Persediaan besar, sedang, dan kecil dengan menggunakan variabel revolving, besaran Uang Persediaan, dan pagu belanja barang. Tidak hanya itu saja, penelitian ini menghasilkan model prediksi total GUP dalam setahun dengan menggunakan variabel yang sama.

Kata kunci: *Uang Persediaan, Analitika, Data.*

A. Pendahuluan

Uang Persediaan (UP) adalah uang muka kerja dalam jumlah tertentu yang diberikan kepada Bendahara Pengeluaran untuk membiayai kegiatan operasional sehari-hari Satuan Kerja (Satker) atau membiayai pengeluaran yang menurut sifat dan tujuannya tidak mungkin dilakukan melalui mekanisme pembayaran langsung (Peraturan Menteri Keuangan Nomor 190 Tahun 2012). UP digunakan untuk keperluan membiayai kegiatan operasional sehari-hari Satker dan membiayai pengeluaran kepada 1 (satu) penerima/penyedia barang/jasa dapat melebihi Rp.50.000.000,- (lima puluh juta rupiah), UP digunakan untuk pengeluaran Belanja Barang, Belanja Modal; dan Belanja Lain-lain. UP merupakan uang muka kerja dari Kuasa BUN kepada Bendahara Pengeluaran yang dapat dimintakan penggantian (revolving) dan telah dipergunakan paling sedikit 50% (lima puluh persen).

Data telah menjadi bagian dari modernitas. Hal ini terlihat dari penggunaan teknologi dalam pemanfaatan data digital yang telah dilakukan pada berbagai hal dan seiring dengan berkembangnya teknologi secara terus menerus, pembuatan konten baru, konektivitas, dan *software* dan infrastruktur sebagaimana dituliskan Avital, Bjorn-Andersen, dan Jeztek (2014). *Big Data* merupakan kumpulan data dengan jumlah yang besar yang dapat merefleksikan transformasi yang substantial dari akumulasi kuantitas yang sederhana menuju perubahan pada kualitas. Kemunculan dan perkembangan dari *big data* ini memberikan dorongan untuk pengembangan *data science* (Cheng X. et al, 2021). *Data Science* adalah teori dasar dan metodologi yang berhubungan untuk memahami makna deretan data dengan menggunakan permodelan, analisis, komputasi, dan pengetahuan untuk mengubah data menjadi informasi, informasi menjadi pengetahuan, pengetahuan menjadi pengambilan keputusan dan memahami kesadaran (*cognition*) dan penyalahgunaan (*manipulation*) pada kondisi nyata.

Data science adalah proses untuk mendestilasi atau mengekstraksi atau menggali insights dari data. Insights dapat berupa informasi penting maupun model-model yang dibuat dari data yang akan bermanfaat dalam mengambil keputusan (UNPAR, 2020). Analitik, sebagai konsep multidisiplin, didefinisikan sebagai cara memperoleh data dari berbagai sumber, memproses untuk memperoleh pola dari informasi dan pengetahuan dan membagikan hasilnya kepada pemangku kepentingan (Selis dan Soltanpoor, 2016). Berbagai penelitian dilakukan untuk mendapatkan metode yang efektif untuk

mengumpulkan, melaporkan, memperoses, memahami, dan mengambil pengetahuan dari *Big Data*. Menurut Selis dan Sultanpoor (2016), *Data analytic* dapat dikelompokkan menjadi tiga kategori, yaitu:

1. *Descriptive Analytic*, yang merupakan bentuk "ringkasan data" dan laporan masa lalu. kategori ini menjawab pertanyaan, "Apa yang telah terjadi" dan mengambil informasi dari data mentah. Terdapat juga kelanjutan dari *descriptive analytic* yaitu "*diagnostic analytic*" yang melaporkan hal yang telah berlalu serta juga mencoba menjawab pertanyaan seperti "Mengapa ini terjadi?". Hal ini membantu organisasi untuk memahami penyebab dari kejadian yang telah berlalu. *Diagnostic analytic* juga membantu untuk memberikan pengetahuan atas hubungan dari berbagai jenis data.
2. *Predictive Analytic*, merupakan bentuk "ramalan atau perhitungan" dan menggabungkan keluaran dari *descriptive analytics* dan juga algoritma *machine learning* serta teknis simulasi untuk membuat model yang memprediksi masa depan dengan akurat. Kategori ini menjawab pertanyaan "Apa yang akan terjadi" dan "Mengapa ini terjadi" di masa depan. *Predictive analytic* membantu untuk mengidentifikasi kesempatan di masa depan dan risiko yang mungkin terjadi dengan membedakan pola yang spesifik dari histori data. Semakin banyak data yang digunakan semakin banyak model yang tervalidasi sehingga semakin akurat prediksi tersebut. Output dari *Predictive analytic* adalah berbagai prediksi dan berekuivalen dengan nilai probabilitasnya.
3. *Prescriptive Analytics*; merupakan bentuk "rekomendasi atau petunjuk" yang memberikan keputusan yang adaptif, otomatis, tidak bergantung waktu, dan optimal. Tujuan dari analitik ini adalah untuk memberikan nilai melalui strategi dan keputusan operasional yang lebih baik. Secara umum, analitik ini adalah *predictive analytics* yang menentukan satu atau lebih tindakan dan menunjukkan dampak dari tiap tindakan tersebut. *Prescriptive analytic* menjawab pertanyaan "Apa yang harus dilakukan?" dan "Mengapa harus melakukan hal tersebut". Analitik ini mempertimbangkan keluaran dari *predictive analytic* yang memenuhi aturan dan kendala untuk menghasilkan tindakan dan keputusan yang optimal. Dengan kata lain, dibutuhkan model prediktif yang mendalam dan dapat ditindaklanjuti serta umpan balik dari yang dikumpulkan dari tindakan dan rekomendasi keputusan yang optimal tersebut untuk membantu pengambil keputusan dalam mencapai dampak yang diharapkan. Sistem *Prescriptive analytics* secara umum memiliki dua hal yang penting, yaitu memberikan dampak yang dapat ditindaklanjuti untuk menghasilkan keputusan yang dapat dipahami tindaklanjutnya dan membantu mekanisme umpan balik dalam melacak rekomendasi yang disarankan serta kejadian yang belum pernah terjadi sebelumnya.

Menurut Runkler, pengerjaan analisis data dapat dibagi menjadi beberapa fase. Data dinilai dan dipilih, dibersihkan dan disaring, divisualisasikan dan dianalisis, dan hasil analisis yang ada kemudian ditafsirkan dan dievaluasi. Seorang Peneliti *Data Science* haruslah dapat melihat permasalahan dari perspektif data (Fawcett, T., Foster, dan Provost, 2013). Terdapat struktur fundamental, cara berpikir dengan *data-analytic* dan prinsip dasar yang harus dipahami. Setidaknya terdapat empat pilar dalam pengembangan aktivitas kompetensi data analisis meliputi kegiatan pencarian dan pengumpulan berbagai sumber data dalam topik tertentu; merumuskan data yang hanya akan digunakan, melakukan persiapan untuk analisis, distribusi dan penyimpanan data; analisis untuk membuat pemahaman terkait dengan makna bentuk data, merancang model dan menciptakan grafis dan pemetaan dari hasil analisis; penelitian terhadap tipe visualisasi, menemukan narasi yang mendasari data dan mendeskripsikan/ menceritakan secara visual (Data Besar, Data Analisis, dan Pengembangan Kompetensi Pustakawan).

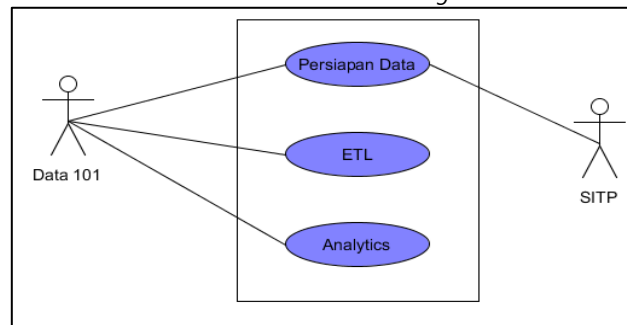
Berdasarkan Neraca Bendahara Umum Negara Kas Bendahara Pengeluaran sebesar Rp731.061.274.902 per 31 Desember 2020 mengalami kenaikan dibandingkan dengan Rp234.188.766.627 per 31 Desember 2019 (Laporan Keuangan Pemerintah Pusat Tahun Anggaran 2020, 2021). Besarnya UP yang dipegang oleh Bendahara menunjukkan adanya potensi *cash idle* pada akhir tahun yang belum disetorkan ke kas Negara. Mengutip kajian (Saad, 2021) Pengeluaran Negara dengan menggunakan mekanisme pembayaran langsung apabila dihubungkan dengan manajemen kas tentunya tidak mengakibatkan terjadinya uang yang mengendap pada Bendahara Pengeluaran. Tapi tidak halnya mekanisme pembayaran UP, banyak terdapat uang yang mengendap dengan diberlakukannya sistem pembayaran ini jika kita amati dari peraturan yang berlaku. Adapun tujuan utama dari manajemen kas itu sendiri yang antara lain adalah mengurangi terjadinya uang yang mengendap pada Bendahara Pengeluaran.

Kajian (Sardjasasmita, 2020) menuliskan beberapa faktor utama yang menyebabkan terjadinya *idle cash* UP pada Satker di seluruh Indonesia. Pertama untuk besaran UP yang selama ini diajukan Satker relatif masih besar dan hanya berdasarkan perkiraan kasar jumlah UP yang mereka butuhkan. Kedua, keterlambatan pengajuan GUP turut disebabkan oleh banyaknya Satker yang menunggu pengajuan GUP hingga mendekati jumlah akumulasi 90-100% sehingga berdampak pada proses revolving UP yang lambat (melebihi 1 bulan). Ketiga adalah Koordinasi dan pengawasan oleh KPPN yang perlu diperketat.

Kajian (Utomo, 2020) pada akhir 2020 menunjukan penyebab dari menurunnya tingkat kepatuhan Satker dalam pertanggungjawaban UP antara lain adanya perubahan sistem pencairan dana yang semula dilakukan dengan tatap muka langsung ke KPPN berubah menjadi sistem daring melalui satu Aplikasi, yaitu E-SPM; kebijakan pemerintah yang mewajibkan seluruh instansi pemerintah menerapkan sistem kerja baru, dimana seluruh atau sebagian pegawai melaksanakan Work From Home (WFH) juga berdampak pada pelaksanaan anggaran di Satker; banyak kegiatan Satker yang tidak berjalan sebagaimana mestinya, kegiatan tersebut telah tergantikan dengan sarana pertemuan daring. Adanya revisi akun belanja Covid-19 dan pemotongan anggaran pada Kementerian Negara/Lembaga juga menjadi penyebab turunnya tingkat kepatuhan Satker.

Manfaat yang diharapkan dari pelaksanaan penelitian ini adalah menjadi bahan rujukan bagi berbagai pihak yang terlibat dalam berbagai proses pengelolaan UP sesuai dengan peran yang dijalankan oleh para pihak.

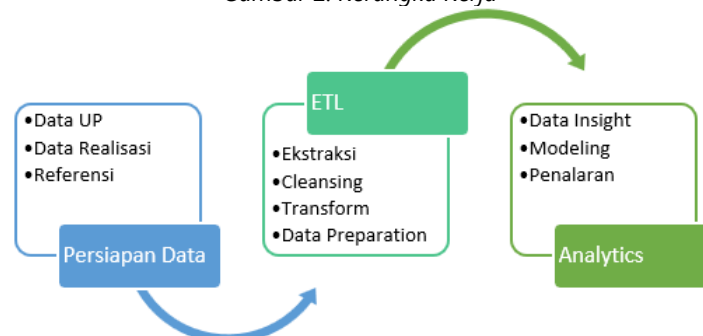
Gambar 1. Use Case Analitika Uang Persediaan



Use case diatas digunakan sebagai batasan dalam mengolah analitika data UP Satker pengguna APBN yang menggambarkan hubungan antara aktor yang terlibat dan aktivitas yang dilakukan dalam menggunakan sebuah sistem. Aktivitas dimulai dengan persiapan data dilanjutkan dengan proses *Extract, Transform* dan *Load* (ETL) data dan diakhiri dengan proses data analytics dan penalaran hasil.

Kerangka kerja yang akan dilakukan adalah sebagai berikut: Data101 akan melakukan proses persiapan data meliputi data permintaan UP, Data revolving UP, Data Realisasi dan beberapa referensi terkait seperti referensi bagian anggaran, Satker, akun, lokasi. Proses dilanjutkan dengan ETL dimana pada tahapan ini akan dilakukan proses extract data yang telah disediakan oleh panitia dilanjutkan dengan *cleansing data* untuk memastikan konten data sesuai dengan kebutuhan dalam proses analitika dilanjutkan dengan proses transform untuk memastikan data siap digunakan dan di load pada sistem untuk disimpan dan digunakan untuk proses analitika. Proses analitika akan mencari *insight* data yang akan bermanfaat untuk pemodelan analitika data baik *descriptive* maupun *predictive* dan melakukan penalaran hasil dari insight dan modeling yang dihasilkan agar dapat dimanfaatkan kedepannya sebagaimana digambarkan pada gambar berikut:

Gambar 2. Kerangka Kerja



B. Metodologi Penelitian

Dalam melakukan penelitian ini digunakan pendekatan *Design Science Research Methodology (DSRM)* adalah salah satu kerangka kerja DSR yang mencakup aktivitas berikut ini:

1. Identifikasi masalah dan motivasi, di mana peneliti perlu menyajikan masalah dunia nyata dan urgensi penyelesaian masalahnya;
2. Menentukan tujuan dan solusi, di mana peneliti perlu menyajikan solusi baru untuk masalah tersebut dan menjelaskan kemanjurannya;
3. Perancangan dan pengembangan, di mana peneliti menerapkan teori dalam solusi baru yang dibuatnya;
4. Demonstrasi, di mana peneliti menunjukkan cara menggunakan artefak untuk menyelesaikan masalah;
5. Evaluasi, di mana peneliti mengamati efektivitas artefak dalam menyelesaikan masalah dan membuat keputusan perbaikan; dan
6. Komunikasi, di mana peneliti menginformasikan artefaknya kepada audien dengan menggunakan sistematika tertentu.

Penggunaan DSRM pada penelitian ini dipilih untuk menghasikan model dan artefak yang berkaitan dengan UP. Lebih lanjut model dan artefak yang digunakan untuk menyelesaikan permasalahan yang ada sesuai dengan ketersediaan data dan kemampuan dalam menggunakan teknologi yang digunakan.

Langkah-langkah yang dilakukan pada penelitian ini untuk menghasilkan analitika data UP Satker pengguna APBN dengan menggunakan teknologi yang 'free' dan dapat digunakan oleh khalayak umum adalah sebagai berikut:

1. Data yang diperlukan dalam format *excel* dan *.csv* akan di *extract* dan di *cleansing* pada *database MySQL/ maria db* dengan pertimbangan besarnya ukuran file dan banyaknya baris transaksi.
2. proses *transform* menentukan tipe data yang akan digunakan untuk mengevaluasi apakah baris data yang diberikan telah konsisten dari baris awal sampai dengan akhir? apakah ditemukan data yang anomali sehingga dapat mempengaruhi hasil analitika? Untuk memudahkan proses ini akan digunakan *Toad for MySQL*.
3. langkah berikutnya adalah menggunakan *Power BI Desktop* untuk proses *transform* dan *load* dan selanjutnya melakukan analisa dan visualisasi data. Untuk menghubungkan koneksi data antara *database MySQL/ maria db* dengan *Power BI Desktop* harus dilakukan instalasi *MySQL Connector* terlebih dahulu. Untuk keperluan analitika data akan digunakan kode bahasa *python* dengan menggunakan beberapa *library* seperti *numpy*, *pandas* dan *Scikit-Learn* serta *matplotlib*. Untuk menjalankan *script Python* dengan menggunakan *Anaconda3* yang merupakan paket distribusi *Python* untuk keperluan pemrograman data science.

Analisa yang digunakan pada penelitian ini adalah analisa deskriptif dan kuantitatif dan didukung oleh visualisasi data.

C. Hasil Penelitian dan Pembahasan

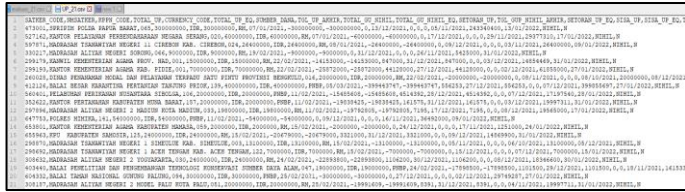
Pada bagian ini berisi tahapan yang dilakukan untuk melakukan analitika data UP Satker pengguna APBN, langkah-langkah yang dilakukan berdasarkan *use case* yang telah disampaikan pada bagian sebelumnya sebagai berikut:

1. Melakukan identifikasi data
2. Melakukan proses *Extract, Transform* dan *Load*
3. Mempersiapkan script *Python* dan *Power BI Desktop*
4. Analitika data UP satuan kerja pengguna APBN

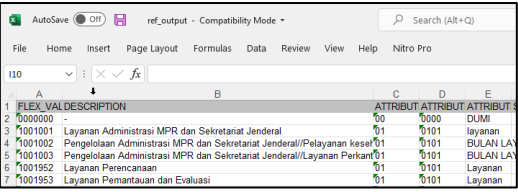
a. Melakukan identifikasi data

Proses identifikasi data meliputi data permintaan UP, Data *revolving* UP, Data Realisasi dan beberapa referensi terkait seperti referensi bagian anggaran, Satker, akun, lokasi. Data Nasional yang digunakan adalah data Tahun Anggaran 2021 sedangkan untuk data pembuatan artefak akan menggunakan data KPPN Sumbawa Besar yang diperoleh dari OMSPAN.

Gambar 3. (a) Data Permintaan UP, Data Revolving UP, dan Data Realisasi UP TA 2021 Satker Lingkup KPPN Sumbawa Besar; (b) Referensi Terkait



(a)



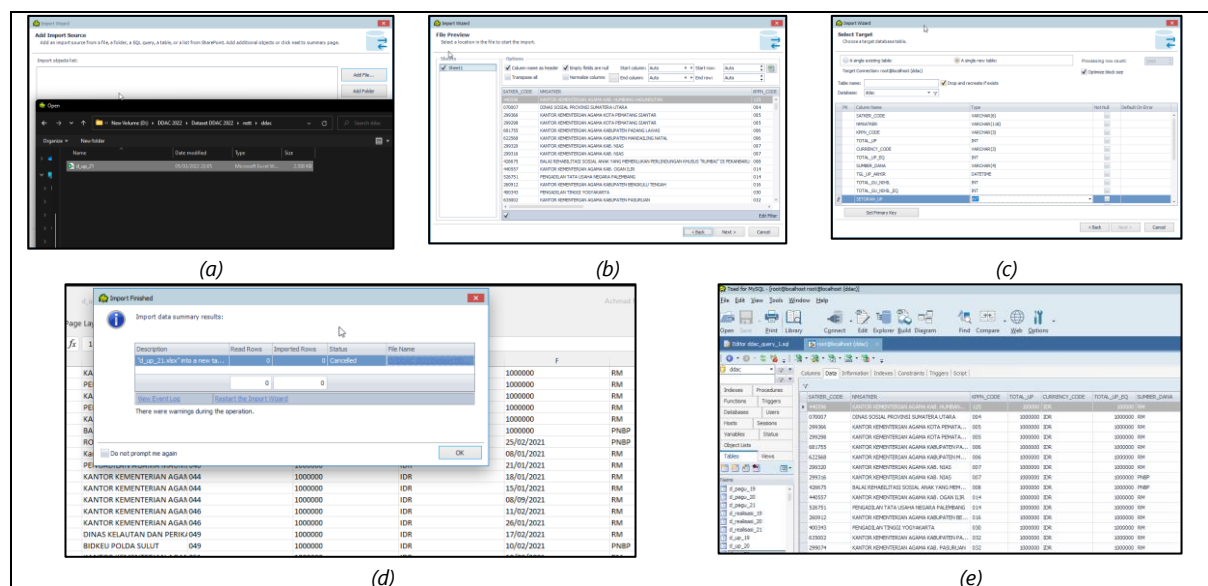
(b)

Sumber: OM SPAN, 2021

b. Melakukan proses *Extract, Transform dan Load*

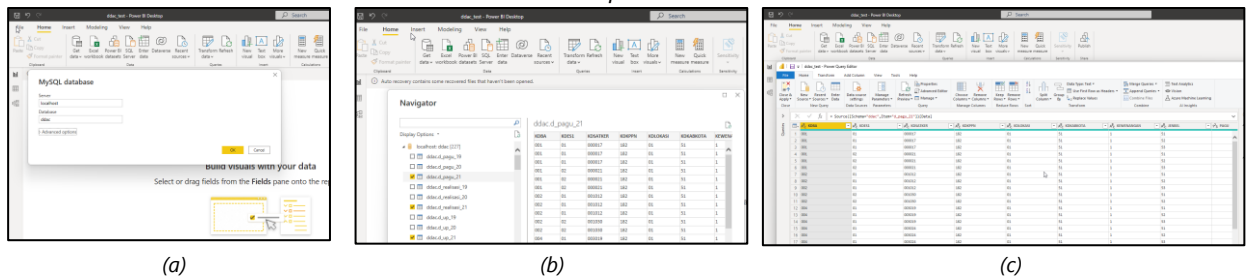
Pada proses ini data diatas dilakukan proses *extract* dan *cleansing data* dengan menggunakan Toad for MySQL dan data disimpan pada database MySQL. Proses ekstraksi yang dilakukan dengan melakukan file data yang akan dilakukan proses upload file data dan referensi yang akan diunggah dan preview data kemudian dilanjutkan dengan proses *mapping* struktur data. Jika proses *export* data mengalami kegagalan maka akan muncul *error* kesalahan sebagai contoh berikut karena ada kolom yang bergeser yang seharusnya numeric terisi dengan date, data yang salah akan disimpan terpisah dan akan dilakukan perbaikan data awal sebelum dilakukan proses *extract* ulang. Data yang sudah berhasil di *extraxct* akan muncul pada database MySQL dengan menggunakan Toad for MySQL.

Gambar 4. (a) Upload File Data dan Referensi; (b) Preview Data; (c) Mapping Struktur Data; (d) Error Data; dan (e) Data yang Berhasil Extract



Setelah data tersimpan pada database MySQL, untuk dapat digunakan pada Power BI Desktop dilakukan proses transform dan load pada Power BI. Pertama kita hubungkan *Power BI* dan *MySQL* setelahnya kita pilih tabel apa saja yang akan digunakan untuk proses analitika dan visualisasi yang akan digunakan. Proses berikutnya adalah proses transform proses *export* data yang telah dilakukan pada database MySQL akan mempersingkat proses transform pada *Power BI desktop*. Setelah proses transform selesai dilakukan apply dan data akan dilakukan load.

Gambar 5. (a) Menghubungkan Power BI dan MySQL; (b) Memilih tabel apa saja yang akan digunakan ; (c) Proses Transform



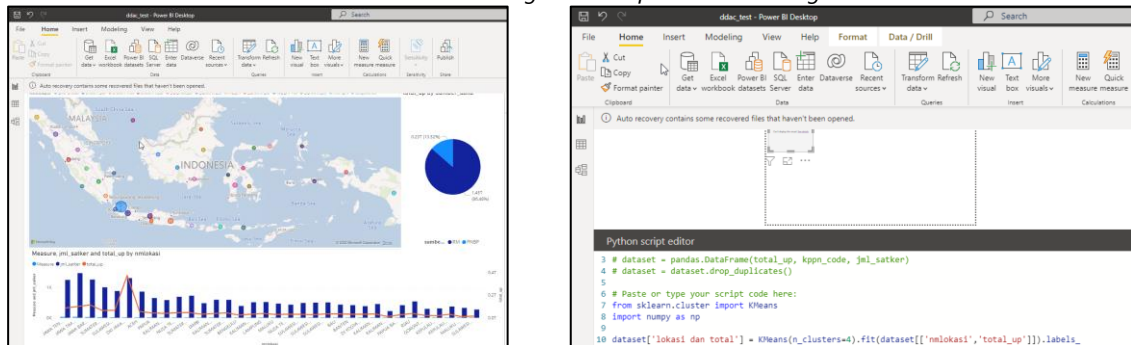
c. Mempersiapkan script Python dan Power BI Desktop Analitika Data

Proses ETL yang telah selesai dilakukan dilanjutkan dengan persiapan script Python dan visualisasi data. Visualisasi dilakukan sebagai *Descriptive Analytic* untuk membantu manajemen dalam memahami uang persediaan yang telah terjadi dan langkah apa yang harus dilakukan kedepannya. Kemudian untuk *Predictive Analytic* akan menggunakan Algoritma supervised learning (yang) membutuhkan data label atau kelas, sedangkan pada algoritma unsupervised learning tidak membutuhkan data label (Kurniawati, 2021).

Algoritma *unsupervised learning* yang digunakan adalah K-Mean Clustering yang bertujuan untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel K. Variabel K sendiri adalah jumlah pengelompokan yang diinginkan (Kurniawati, G. N., 2021). Algoritma *unsupervised learning* lebih bebas dalam proses eksplorasi data karena tidak memiliki data label dan bisa mencari karakteristik data yang tersembunyi. Algoritma ini menggunakan titik data sebagai referensi untuk menemukan struktur dan pola yang ada di dalam data set.

Setelah proses eksplorasi selesai dilakukan akan dilanjutkan dengan Algoritma supervised learning dengan menggunakan *Linear Regression* adalah metode yang mencoba untuk memodelkan hubungan antar variabel dengan menyesuaikan persamaan linear (Ramadhani, R. D., 2019). Satu variabel sebagai dependent variabel atau target dan variabel yang lain sebagai independent variabel (*explanatory variabel*).

Gambar 6. Algoritma Supervised Learning



d. Analitika data uang persediaan satuan kerja pengguna APBN

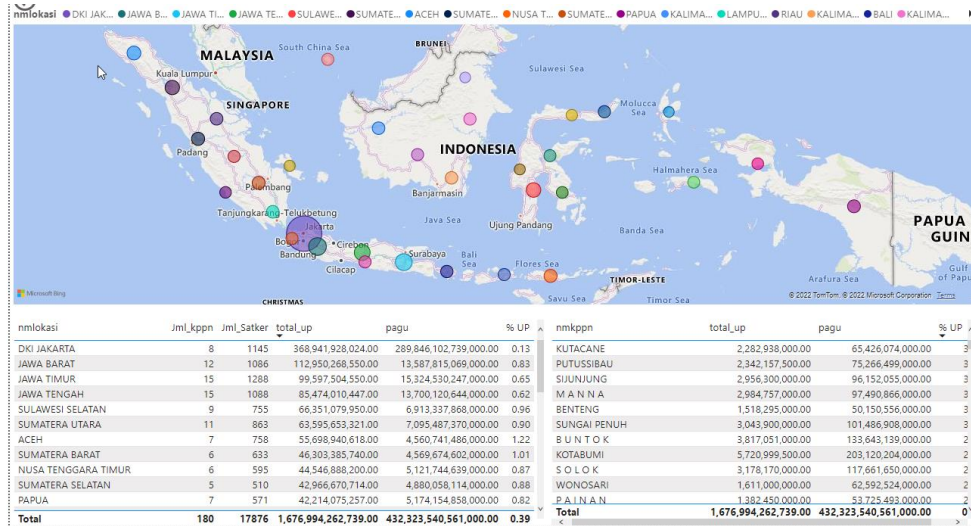
(1) *Descriptive Analytic*

Descriptive Analytic UP Satker pengguna APBN dilakukan dengan visualisasi dengan menggunakan Power BI Desktop

(a) Analisis berdasarkan besaran UP

Hasil visualisasi berdasarkan besaran UP dengan menggunakan *Power BI Desktop* sebagai berikut:

Gambar 5. Besaran UP di Seluruh Indonesia



Sumber: Dataset UP dan Pagu 2021, 2021

Hasil visualisasi menunjukkan secara nasional besaran UP masih didominasi di Pulau Jawa tepatnya Provinsi DKI Jakarta diikuti berturut-turut dengan Jawa Barat, Jawa Timur dan Jawa Tengah. Sedangkan terkecil adalah Maluku Utara dan Kalimantan Utara.

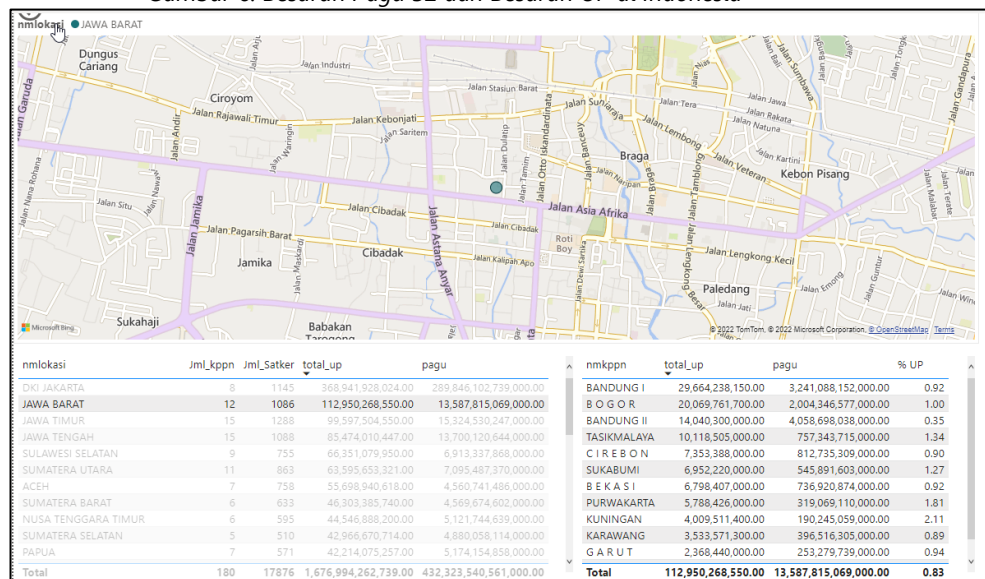
Berdasarkan data pagu 52 dan data UP dapat dibandingkan satu lokasi dengan lokasi yang lain, seperti contohnya Jawa Barat dan Jawa Tengah, walaupun pagu 52 Jawa Tengah (Rp13,7 T) lebih besar sedikit dari Jawa Barat (Rp13,58T) tetapi total UP yang diberikan Jawa Tengah (Rp.85,47 M) lebih kecil dari Jawa Tengah (Rp112,95 M). Jika dibandingkan dengan 10 lokasi lain diluar DKI Jakarta, Jawa Tengah merupakan Kanwil yang paling rendah menyalurkan UP yaitu 0,62%.

Dari informasi yang ada Direktorat Teknis dapat melakukan penelitian lebih lanjut terkait penyesuaian kebijakan penyaluran UP kepada Satker, misalkan maksimal nilai UP yang dapat disalurkan diawal tahun berapa persen dari total pagu 52 sesuai dengan besaran pagu yang dimiliki oleh masing-masing Kanwil, keuntungan lain *idle cash* di level nasional dapat lebih ditekan. Kanwil Jawa Tengah dapat dijadikan narasumber bagaimana mengelola pemberian UP yang baik untuk Satker yang dapat dicontoh oleh kanwil lain.

(b) Analisis berdasarkan besaran Pagu 52, UP, Jumlah KPPN dan Jumlah Satker

Hasil visualisasi berdasarkan besaran Pagu 52, UP, Jumlah KPPN dan Jumlah Satker dengan menggunakan *Power BI Desktop* sebagai berikut:

Gambar 6. Besaran Pagu 52 dan Besaran UP di Indonesia



Sumber: Dataset UP dan Pagu 2021, 2021

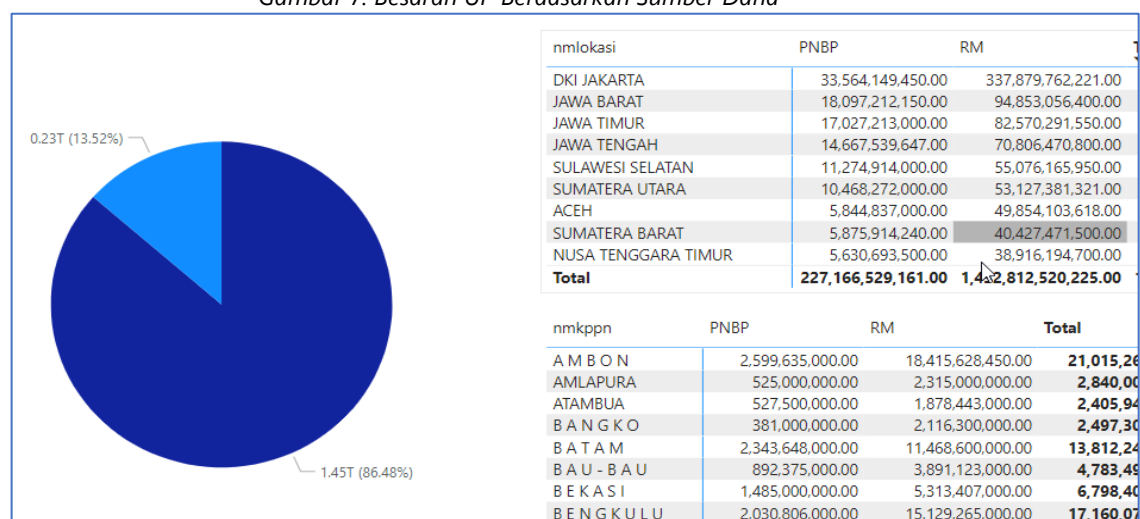
Hasil dengan menggunakan tabel dibawah peta akan diperoleh gambaran yang lebih detail bila dibandingkan dengan analisa sebelumnya, untuk proses lebih detail akan fokus pada Lokasi Jawa Barat, tampilan dashboard peta dan tabel detail KPPN akan muncul hanya pada wilayah kerja Kanwil Jawa Barat. Analisa lebih lanjut pada KPPN manakah yang menyalurkan UP RM kepada Satker dapat dilihat lebih detail seperti gambar diatas. Bandung II walaupun memiliki Pagu 52 terbesar tetapi menyalurkan UP paling rendah yaitu 0.35% sedangkan untuk persentase terbesar adalah Kuningan (2.11%), Purwakarta (1.81%), Tasikmalaya (1.34%) dan Sukabumi (1.27%).

Berdasarkan informasi yang ada Kanwil dapat lebih melakukan pembinaan dan supervisi kepada KPPN di lingkup kerja sehingga Kanwil dan KPPN dapat meningkatkan nilai IKPA khususnya terkait dengan Pengelolaan UP dan TUP. Disisi lain Satker dengan pengelolaan UP yang sesuai dengan kebutuhan dapat memudahkan Satker untuk segera melakukan revolving UP. KPPN Bandung II dapat dijadikan narasumber bagaimana mengelola UP dan indikator apa saja yang perlu diperhatikan dalam pemberian UP di awal tahun kepada Satker yang dapat dicontoh oleh KPPN di lingkup Kanwil Jawa Barat. Demikian halnya kepada KPPN yang secara nasional baik dalam pengelolaan UP dan pemberian UP di awal tahun, dapat dijadikan narasumber sesuai dengan pengelompokkan KPPN tipe A1 Provinsi, A1 Non Provinsi, dan A2.

(c) Analisis berdasarkan sumber dana

Hasil visualisasi berdasarkan sumber dana dengan menggunakan *Power BI Desktop* sebagai berikut:

Gambar 7. Besaran UP Berdasarkan Sumber Dana



Sumber: Dataset UP dan Pagu 2021, 2021

Hasil visualisasi menunjukkan besaran UP berdasarkan sumber dana, besaran porsi Uang Persediaan dengan Rupiah Murni adalah 86,4% dan PNBP adalah 13,6%, bila data di detailkan terlihat porsi PNBP berbanding lurus dengan besaran UP yang disalurkan di setiap lokasi. Demikian hal nya dengan KPPN lingkup kerja Kanwil akan terlihat besaran UP dengan sumber dana PNBP maupun RM.

(2) Predictive Analytic

Predictive Analytic dijalankan dengan menggunakan Anaconda3 menjalankan *script Python*.

(a) K-Mean Clustering

K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain (Ohorella, 2019). Pada pembahasan ini akan digunakan data KPPN Sumbawa Besar yang diolah dari OMSPAN terkait dengan jumlah revolving UP, besaran UP, total GUP, Pagu 52 dan Realisasi 52. Langkah yang dilakukan pada proses ini adalah sebagai berikut

- Pertama kali yang dilakukan adalah import semua library yang dibutuhkan seperti pandas, numpy, matplotlib dan sklearn dan dilanjutkan dengan membaca data .csv pada direktori

data disimpan dan untuk memastikan data sudah dibaca kita dapat melakukan statistik deskriptif atas data dimaksud

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

df = pd.read_csv('D:\\DOAC 2022\\Dataset DOAC 2022\\netf\\Data101.csv')
df.describe()
```

	Satker	Revolving	UP	GUP	Pagu52	Real52
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	497359.254902	21.176471	43.470588	936.254902	2696.078431	2583.627451
std	176016.955401	8.744612	69.203859	2051.793887	4351.902328	4310.550127
min	8800.000000	3.000000	5.000000	17.000000	95.000000	94.000000
25%	417676.500000	16.000000	15.000000	209.000000	651.500000	641.000000
50%	553612.000000	21.000000	21.000000	360.000000	1158.000000	1151.000000
75%	656875.000000	24.500000	46.500000	932.000000	2604.500000	2572.000000
max	690154.000000	49.000000	400.000000	12686.000000	21672.000000	21601.000000

- Kemudian lakukan pemilihan variabel yang akan digunakan untuk proses analitika data yang akan digunakan yaitu jumlah revolving UP selama 1 tahun (Revolving), kemudian besaran UP yang diterima Satker di awal tahun (UP) dan besaran Pagu belanja barang Satker bersangkutan (Pagu52), kemudian lanjutkan dengan melihat beberapa baris awal dari data

```
data = df[['Revolving', 'UP', 'Pagu52']]
data.head()
```

	Revolving	UP	Pagu52
0	43	30	1991
1	37	30	1572
2	14	14	146
3	11	15	596
4	3	15	3993

- Siapkan satu variabel yang akan menampung hasil proses pengelompokan yaitu 'y' dan dilanjutkan dengan melakukan proses Kmeans yang akan menuliskan hasil pengelompokan pada variabel yang telah disiapkan. Hasil proses menghasilkan kelompok rendah (0), tinggi (1) dan sedang (2).

```
X = data.values
k_means_optimum = KMeans(n_clusters = 3, init = 'k-means++', random_state=48)
y = k_means_optimum.fit_predict(X)
print(y)
```

```
[0 0 0 0 2 0 2 0 0 0 2 0 0 0 0 2 0 0 0 0 2 0 1 0 0 0 2 0 0 0 0 1 2 0 0 0
 2 0 0 0 0 0 0 1 2 0 0 0 0 0]
```

- Hasil Kmeans dari 3 variabel yang telah ditetapkan dan satu variabel y sebagai hasil pengelompokan akan menghasilkan pengelompokan data sebagai berikut

Kelompok Rendah	39 Satker
Kelompok Sedang	9 Satker
Kelompok Tinggi	3 Satker

```
data['cluster'] = y
data1 = data[data.cluster==0]
data2 = data[data.cluster==1]
data3 = data[data.cluster==2]
data['cluster'].value_counts()
```

0	39
2	9
1	3

Name: cluster, dtype: int64

- Untuk melakukan visualisasi pengelompokan dengan menggunakan variabel yang telah ditetapkan, pilih 3 warna yang akan merepresentasikan hasil kelompok yaitu kelompok rendah (merah), kelompok sedang (biru) dan kelompok tinggi (hijau). Selain itu kita akan memperoleh titik pusat kelompok dengan menggunakan warna indigo. Berikan judul Visualisasi yaitu 'Kmeans'

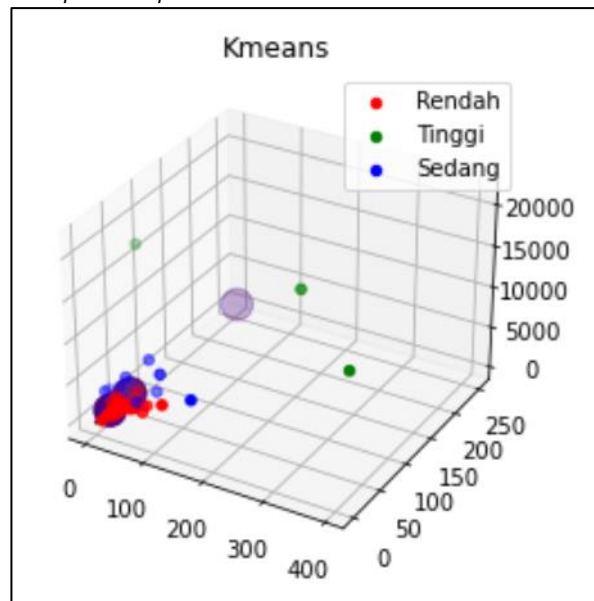
```

kplot = plt.axes(projection='3d')
xline = np.linspace(0, 15, 10000)
yline = np.linspace(0, 15, 10000)
zline = np.linspace(0, 15, 10000)
kplot.plot3D(xline, yline, zline, 'black')
# Data for three-dimensional scattered points
kplot.scatter3D(data1.UP, data1.Revolving, data1.Pagu52, c='red', label = 'Rendah')
kplot.scatter3D(data2.UP, data2.Revolving, data2.Pagu52, c = 'green', label = 'Tinggi')
kplot.scatter3D(data3.UP, data3.Revolving, data3.Pagu52, c = 'blue', label = 'Sedang')
plt.scatter(k_means_optimum.cluster_centers[:,0], k_means_optimum.cluster_centers[:,1], color = 'indigo', s = 200)
plt.legend()
plt.title("Kmeans")
plt.show()

```

- Dibawah ini merupakan visualisasi KMeans hasil cluster dari 3 variabel yaitu jumlah frekuensi revolving, jumlah UP dan jumlah Pagu 52, yang di representasikan dengan warna hijau untuk kelompok tinggi (3 Satker), warna biru untuk kelompok sedang (9 Satker) dan warna merah untuk kelompok rendah (39 Satker)

Grafik 1. Grafik K-Means UP di KPPN Sumbawa Besar



Sumber: OM SPAN, data diolah

Berdasarkan hasil visualisasi, dapat diperoleh informasi terdapat hubungan antara variabel besaran UP, *revolving* UP, dan besaran pagu 52. Namun demikian terdapat hasil mencolok/*outlier* yakni tiga Satker tergolong kelompok tinggi (titik hijau). Hal ini menandakan 3 Satker tersebut memiliki variabel lain yang memengaruhi pengelompokan selain variabel *revolving* UP, besaran UP dan besaran pagu 52.

(b) Linear Regression

Pada bagian ini akan dilakukan Linear Regression yang akan memodelkan hubungan antar variabel dengan menyesuaikan persamaan linear. Satu variabel sebagai dependent variabel atau target dan variabel yang lain sebagai independent variabel (*explanatory variabel*). Pada pembahasan ini akan digunakan data KPPN Sumbawa Besar yang diolah dari OMSPAN terkait dengan jumlah revolving UP, besaran UP, total GUP, Pagu 52 dan Realisasi 52. Langkah yang dilakukan pada proses ini adalah sebagai berikut:

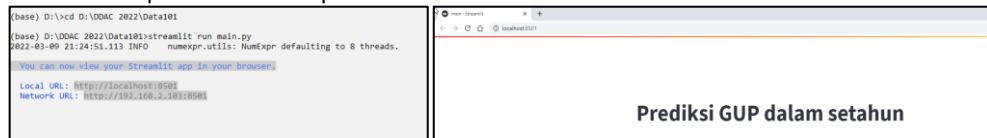
- Langkah pertama yang dilakukan adalah membuka visual editor melalui anaconda, lalu buat folder yang akan digunakan untuk menyimpan script yang dibuat. Kemudian buatlah file main.py kemudian import semua library yang dibutuhkan seperti pandas, numpy, matplotlib, sklearn, seaborn dan streamlit. Lanjutkan dengan memasukkan Judul dari halaman aplikasi

```

File Edit Selection View Go Run Terminal Help
EXPLORER
  DATA01
    Data101.csv
    main.py
main.py
1 import streamlit as st
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 from sklearn.linear_model import LinearRegression
7 from sklearn.model_selection import train_test_split
8 from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
9
10 st.title("Prediksi GUP dalam setahun")

```

- Kemudian melalui membuka command prompt melalui anaconda, lalu arahkan pada directory folder yang telah dibuat pada proses sebelumnya dan lakukan inisiasi streamlit dan akan muncul tampilan streamlit pada web browser



- Lanjutkan dengan membaca data .csv pada direktori data disimpan, kemudian buang kolom yang tidak diperlukan pada proses berikutnya kemudian tampilkan isi data yang sudah siap digunakan

```

12 load_dataset
13 def load_dataset():
14     df = pd.read_csv('D:\ODAC 2022\Data101\Data101.csv')
15     return df
16
17 df = load_dataset()
18 df.drop(['Satker', 'Real152'], axis=1, inplace=True)
19
20 # show dataset
21 st.header("Dataset")
22 st.write(df)

```

Prediksi GUP dalam setahun

Dataset

	Revolving	UP	GUP	Pagu52
0	43	30	1259	1991
1	37	30	1017	1572
2	14	14	131	146
3	11	15	111	596

- Langkah berikutnya ada menentukan besaran parameter test_size yang berfungsi untuk menentukan ukuran data testing. Dalam contoh di atas, test_size=0.3 berarti data yang akan digunakan sebagai data testing sebesar 30% dari keseluruhan dataset dan random_state merupakan random number generator (RNG) yang akan digunakan untuk menghasilkan angka yang digunakan dalam pengujian. Besarnya test_size dan random_state akan menggunakan slider.

```

24 st.header("Tune Parameters")
25 test_size = st.slider('Test Size', 0.1, 0.5, 0.1)
26 random_state = st.slider('Random State', 0, 200, 1)

```

The screenshot shows the 'Tune Parameters' section with two sliders. The 'Test Size' slider is set to 0.1, and the 'Random State' slider is set to 1.

- Proses dilanjutkan dengan pembagian data training dan data testing dan membuat modelling machine learning yang digunakan untuk melakukan training dan testing

```

28 X = df.drop('GUP', axis=1)
29 y = df['GUP']
30 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=random_state)
31
32 # training
33 def train_model(X_train, y_train):
34     linear_regression = LinearRegression()
35     linear_regression.fit(X_train, y_train)
36     return linear_regression
37
38 # testing
39 def testing_model(model, X_test, y_test):
40     y_pred = model.predict(X_test)
41     return y_pred
42
43 # training model
44 model = train_model(X_train, y_train)
45
46 # testing model
47 y_pred = testing_model(model, X_test, y_test)

```

- Langkah berikutnya melakukan pengujian atas model untuk melihat apakah modelling yang telah dibuat pada proses sebelumnya sudah optimal

```

46 # evaluation
47 def evaluate(y_pred, y_test):
48     mae = mean_absolute_error(y_test, y_pred)
49     rmse = np.sqrt(mean_squared_error(y_test, y_pred))
50     r2 = r2_score(y_test, y_pred)
51     e = ['MAE', 'RMSE', 'R-Squared']
52     eval = pd.DataFrame([mae, rmse, r2], index=e, columns=['Score'])
53     return eval
54
55 # show evaluation
56 st.header('Model Performance')
57 eval = evaluate(y_pred, y_test)
58 st.write(eval)

```

The screenshot shows the 'Tune Parameters' section with sliders for 'Test Size' and 'Random State'. Below it, the 'Model Performance' table is displayed.

	Score
MAE	293.5245
RMSE	526.3418
R-Squared	0.8192

Tulisan (Meiryani, 2021) menjelaskan nilai R-Square dikategorikan kuat jika lebih dari 0,67, moderat jika lebih dari 0,33 tetapi lebih rendah dari 0,67, dan lemah jika lebih dari 0,19 tetapi

lebih rendah dari 0,33. Jika nilai mendekati 1, artinya variabel independen memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variabel dependen. Namun, jika nilai R² semakin kecil, artinya kemampuan variabel – variabel independen dalam menjelaskan variabel dependen cukup terbatas.

- Ketika model yang sudah dibuat optimal maka berikutnya adalah menggunakan prediksi atas model yang telah dibuat, yaitu melakukan prediksi GUP

```
59 st.header("Prediksi Total GUP")
60 Revolving = st.slider('Revolving', 0, 90, 1)
61 UP = st.slider('UP (dalam Juta Rupiah)', 0, 500, 1)
62 Pagu52 = st.slider('Pagu52 (dalam Juta Rupiah)', 0, 25000, 1)
63 # prediksi
64 predictions = model.predict([[Revolving, UP, Pagu52]])
65 if (st.button("Lihat Hasil")):
66     st.header("Prediksi GUP dalam setahun:    Rp.{0:,.000,000}".format(int(predictions)))
```

Dengan menggunakan slider dari variabel yang akan dipilih oleh pengguna yaitu Revolving, UP dan Jumlah besaran Pagu52

Gambar 9. Prediksi Frekuensi Revolving UP, Jumlah UP, dan Pagu 52 pada KPPN Sumbawa Besar



Sumber: OM SPAN, data diolah

Dari prediksi diatas dengan pagu 52 sebesar Rp9,2 Miliar dan persetujuan pemberian UP sebesar Rp70 Juta oleh KPPN serta perkiraan rencana revolving 32 kali dalam setahun maka total GUP setahun yang akan direalisasikan oleh Satker sebesar Rp1,72 Miliar dengan R-Square sebesar 0,81 yang menunjukkan ketiga variabel independen (*revolving* UP, besaran UP dan besaran pagu 52) memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variabel dependen yaitu total GUP setahun yang akan direalisasikan oleh Satker.

D. Kesimpulan dan Saran

Analitika merupakan hal yang diperlukan untuk mengubah data yang ada menjadi informasi, sehingga informasi tersebut dapat digunakan untuk mengambil keputusan, tidak terkecuali dalam implementasi Uang Persediaan pada pemerintahan. Penelitian ini bertujuan untuk memberikan *insight* kepada pihak yang terlibat dalam berbagai proses pengelolaan UP. Penelitian ini menggunakan dua jenis analitika, yaitu *Descriptive Analysis* dan *Predictive Analysis*. *Descriptive Analysis* sendiri dilakukan menggunakan aplikasi tanpa biaya, yaitu *Python* dan *Power BI Desktop*, yang berarti analitika dapat dilakukan oleh siapa saja yang ingin melakukannya. *Predictive Analysis* dilakukan menggunakan analisis *K-Means* dan analisis *Linear Regression*.

Kesimpulan dari analitika yang dilakukan menggunakan *Python* dan *Power BI Desktop* dengan data UP nasional adalah bahwa pemberian UP secara keseluruhan di Indonesia untuk belanja 52 dipandang besar, yaitu senilai Rp1,67 T dari Rp 432,32 T total pagu 52, atau sebesar 0,39%. Besarnya nilai UP yang dikelola Bendahara Pengeluaran dapat mengakibatkan tingginya *idle cash* secara nasional, yang seharusnya dapat digunakan untuk keperluan lain.

Saran penulis adalah agar Direktorat Teknis dapat melakukan tinjauan lebih lanjut mengenai kebijakan pengelolaan UP pada Satker seperti persentase maksimal besaran UP yang dapat diajukan dari total pagu 52. Selain untuk mengurangi *idle cash*, hal ini dapat membantu Kanwil dan KPPN untuk meningkatkan nilai IKPA khususnya Pengelolaan UP dan TUP, hal ini juga memudahkan Satker untuk melakukan *revolving* UP karena UP yang diberikan sesuai dengan kebutuhan Satker tersebut. KPPN

dapat melakukan *Predictive Analysis* untuk mengetahui berapa besar perkiraan realisasi GUP yang mempengaruhi penyerapan anggaran Satker terutama belanja barang.

E. Implikasi dan Keterbatasan

Keterbatasan dalam penelitian ini adalah kurangnya data sampel UP secara nasional, sehingga pada penelitian ini hanya menggunakan data UP yang dikelola oleh KPPN Sumbawa Besar untuk melakukan *Predictive Analysis*. Penggunaan data UP secara nasional dapat meningkatkan validitas prediksi yang lebih akurat. Jenis pagu belanja barang yang digunakan untuk proses penelitian berikutnya perlu didetailkan khususnya akun belanja operasional. Eksplorasi penggunaan variabel dependen lain yang memengaruhi UP perlu dilakukan sehingga menambah variasi variabel yang dapat digunakan pada penelitian berikutnya.

Implikasi untuk penyempurnaan penelitian selanjutnya adalah agar dapat menggunakan data UP secara nasional. Hal ini diperlukan agar tujuan dari penulisan penelitian dapat tercapai dengan lebih komprehensif, yaitu untuk menjadi rujukan dalam proses pengelolaan UP yang lebih baik.

REFERENSI

- Arman, A.A., Cahyana, R., Fitriani L., Mahayana, D, Setiawan, Y. (2021). *Design Science Research Methodology Dalam Topik Sistem Deteksi Ujaran Kebencian Otomatis*. Bandung.
- Avital, M., Bjorn-Andersen, dan N., Jetzek, T.; (2014). Data-Driven Innovation through Open Government Data. *Journal of theoretical and applied electronic commerce research*, 9(2), 15–16. doi:10.4067/S0718-18762014000200008
- Cheng, X., Tang, N., Xu, C., dan Xu, Z. (2021). Data Science: Connotation, Methods, Technologies, And Development. *Data Science and Management*, 1(1), 32–37. doi:10.1016/j.dsm.2021.02.002
- Runkler, T. A. (2016). Computer Science: Models and Algorithms for Intelligent Data Analysis. *Data Analytics*. doi.org/10.1007/978-3-658-14075-5.
- Fakultas Teknologi Informasi dan Sains Universitas Katolik Parahyangan (UNPAR). (2020). *Pengantar Data Science dan Aplikasinya Bagi Pemula*. Bandung.
- Kuangan, K. (2021). Laporan Keuangan Pemerintah Pusat Tahun Anggaran 2020.
- Kuangan, K. (2012). Peraturan Menteri Keuangan Republik Indonesia Nomor 190/PMK.05/2012 tentang Tata Cara Pembayaran dalam Rangka Pelaksanaan Anggaran Pendapatan dan Belanja Negara. Jakarta.
- Kurniawati, G. N. (2021, 4 Februari). *Algoritma Supervised vs Unsupervised Learning, Apa Bedanya?*. <https://www.dqlab.id/algoritma-supervised-vs-unsupervised-learning-apa-bedanya>. Diakses 6 Maret 2022.
- Meiryani. (2021, 12 Agustus). *Memahami Koefisien Determinasi dalam Regresi Linear*. <https://accounting.binus.ac.id/2021/08/12/memahami-koefisien-determinasi-dalam-regresi-linear/>. Diakses pada 7 Maret 2022.
- Narendra, A. P. (2015). Record and Library Journal. *Data Besar, Data Analisis, dan Pengembangan Kompetensi Pustakawan* (volume 1). Salatiga
- Ohorella, N. A. T. (2019, 7 Juli). *Clustering Menggunakan Metode K-Means Cluster*. <https://medium.com/@16611094/clustering-menggunakan-metode-k-means-cluster-di-r-80827b4a48fd>. Diakses pada 7 Maret 2022.
- Provost, Foster; Fawcett, Tom (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51–59. doi:10.1089/big.2013.1508
- Ramadhani, R. D., (2019, 6 Januari). *Memahami K-Mean Clustering pada Machine Learning dengan Python*. <https://medium.com/@16611129/memahami-k-mean-clustering-pada-machine-learning-dengan-python-430323d80868>. Diakses 6 Maret 2022.
- Saad, I. (2021, 13 Desember). *Beberapa Aspek Pengelolaan Uang Persediaan/Tambahan Uang Persediaan Dalam Manajemen Kas*. <https://edjpb.kemenkeu.go.id/forum/search/content-slug/7832-beberapa-aspek-pengelolaan-uang-persediaan-tambahan-uang-persediaan-dalam-manajemen-kas>. Diakses 6 Maret 2022.
- Sardjasasmita, A. (2020, 16 Desember). Pengelolaan Uang Persediaan yang Handal untuk Optimalkan Potensi Dana APBN. <https://edjpb.kemenkeu.go.id/forum/search/content-slug/4427-pengelolaan-uang-persediaan-yang-handal-untuk-optimalkan-potensi-dana-apbn>. Diakses 6 Maret 2022.
- Sellis, T. dan Soltanpoor, R. (2016). Prescriptive Analytics for Big Data. *Databases Theory and Applications*, 245–256. doi:10.1007/978-3-319-46922-5_19
- Taufan, G. (2021, 11 Februari). *Pengenalan Linear Regresi dan Implementasi dengan Python*. <https://gigastaufan-65282.medium.com/pengenalan-linear-regresi-dan-implementasi-dengan-python-afa3dcedec13>. Diakses 6 Maret 2022.
- Utomo, I. S. (2020, 9 Desember). Carut Marut Uang Persediaan di Masa Pandemi Covid-19. <https://edjpb.kemenkeu.go.id/forum/search/content-slug/3416-carut-marut-uang-persediaan-di-masa-pandemi-covid-19> diakses 6 Maret 2022.

LAMPIRAN 1

Script K-Means Cluster

```
# %%
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

df = pd.read_csv('D:\\DDAC 2022\\Dataset DDAC 2022\\nett\\Data101.csv')
df.describe()

# %%
data = df[['Revolving','UP','Pagu52']]
data.head()

# %%
X = data.values

# %%
X = data.values
k_means_optimum = KMeans(n_clusters = 3, init = 'k-means++', random_state=48)
y = k_means_optimum.fit_predict(X)
print(y)

# %%
data['cluster'] = y
data1 = data[data.cluster==0]
data2 = data[data.cluster==1]
data3 = data[data.cluster==2]
data['cluster'].value_counts()

# %%
kplot = plt.axes(projection='3d')
xline = np.linspace(0, 15, 10000)
yline = np.linspace(0, 15, 10000)
zline = np.linspace(0, 15, 10000)
kplot.plot3D(xline, yline, zline, 'black')
# Data for three-dimensional scattered points
kplot.scatter3D(data1.UP,data1.Revolving,data1.Pagu52,c='red',label = 'Rendah')
kplot.scatter3D(data2.UP,data2.Revolving,data2.Pagu52,c = 'green', label = 'Tinggi')
kplot.scatter3D(data3.UP,data3.Revolving,data3.Pagu52,c = 'blue', label = 'Sedang')
plt.scatter(k_means_optimum.cluster_centers_[:,0], k_means_optimum.cluster_centers_[:,1], color = 'indigo', s = 200)
plt.legend()
plt.title("Kmeans")
plt.show()

# %%
data

# %%
data.to_csv('D:\\DDAC 2022\\Dataset DDAC 2022\\nett\\Data101Result.csv')

# %%
```

Lampiran 2
Hasil K-Means Cluster

	Revolving	UP	Pagu52	cluster
0	43	30	1991	0
1	37	30	1572	0
2	14	14	146	0
3	11	15	596	0
4	3	15	3993	2
5	22	18	1236	0
6	23	90	6519	2
7	22	20	843	0
8	24	21	1156	0
9	4	8	200	0
10	16	38	5451	2
11	16	48	1808	0
12	23	60	1937	0
13	20	45	1604	0
14	19	9	860	0
15	15	9	3367	2
16	19	9	1751	0
17	16	6	111	0
18	11	8	319	0
19	32	9	321	0
20	25	90	2639	0
21	34	60	6971	2
22	22	24	655	0
23	18	60	21672	1
24	23	36	1007	0
25	14	21	670	0
26	25	15	648	0
27	26	81	3993	2
28	25	21	877	0
29	17	18	643	0
30	15	9	577	0
31	33	7	451	0
32	49	300	19587	1
33	18	24	3823	2
34	24	30	1571	0
35	27	42	1135	0
36	19	20	538	0
37	15	150	5198	2
38	20	15	2570	0
39	24	15	1519	0
40	30	15	793	0
41	22	15	1039	0
42	21	15	860	0
43	19	5	95	0
44	23	400	13786	1
45	12	60	3058	2
46	21	60	1294	0
47	22	21	679	0
48	14	20	1134	0
49	29	60	1624	0
50	4	6	103	0

Lampiran 3

Script Linear Regression

```
import streamlit as st
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

st.title("Prediksi GUP dalam setahun")

# load dataset
def load_dataset():
    df = pd.read_csv('D:\DDAC 2022\Data101\Data101.csv')
    return df

df = load_dataset()
df.drop(['Satker', 'Real52'], axis=1, inplace=True)

# show dataset
st.header("Dataset")
st.write(df)

st.header("Tune Parameters")
test_size = st.slider('Test Size', 0.1, 0.9, 0.1)
random_state = st.slider('Random State', 0, 200, 1)

X = df.drop('GUP', axis=1)
y = df['GUP']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=random_state)

# training
def train_model(X_train, y_train):
    linear_regression = LinearRegression()
    linear_regression.fit(X_train, y_train)
    return linear_regression

# testing
def testing_model(model, X_test, y_test):
    y_pred = model.predict(X_test)
    return y_pred

# training model
model = train_model(X_train, y_train)

# testing model
y_pred = testing_model(model, X_test, y_test)

# evaluation
def evaluate(y_pred, y_test):
    mae = mean_absolute_error(y_test, y_pred)
    rmse = np.sqrt(mean_squared_error(y_test, y_pred))
    r2 = r2_score(y_test, y_pred)
    e = ['MAE', 'RMSE', 'R-Squared']
    eval = pd.DataFrame([mae, rmse, r2], index=e, columns=['Score'])
    return eval

# show evaluation
st.header('Model Performance')
eval = evaluate(y_pred, y_test)
st.write(eval)

st.header('Prediksi Total GUP')
Revolving = st.slider('Revolving', 0, 90, 1)
UP = st.slider('UP (dalam Juta Rupiah)', 0, 500, 1)
Pagu52 = st.slider('Pagu52 (dalam Juta Rupiah)', 0, 25000, 1)

# prediksi
predictions = model.predict([[Revolving, UP, Pagu52]])
if (st.button("Lihat Hasil")):
    st.header("Prediksi GUP dalam setahun: Rp.{0:,}000,000".format(int(predictions)))
```