

Nama Kursus : Microsoft Azure AI Fundamentals
Mentee : Amelia Angraini M
Universitas : Universitas Singaperbangsa Karawang
Mentor : Agus Suparno
Pertemuan : Asinkron Kedua

Anda bekerja pada sebuah taman nasional konservasi komodo, sebagai karyawan dibagian IT Anda bertanggung jawab terhadap layanan IT yang ada di Instansi tersebut. Atasan Anda meminta Anda untuk meneliti komodo-komodo yang ada di taman nasional tersebut, untuk mengelompokkan komodo-komodo tersebut berdasarkan kemiripan dan proporsi tubuh mereka. Langkah-langkah yang harus Anda lakukan adalah:

1. Memilih model machine learning apa yang cocok untuk permasalahan tersebut? Model Regression, Classification atau Clustering? Kenapa Anda memilih salah satu model tersebut? Jelaskan alasannya? Bandingkan perbedaan dari ketiga model machine learning tersebut!

Jawaban: Saya akan memilih model Clustering. Karena model clustering adalah Teknik unsupervised machine learning tanpa pengawasan yang digunakan untuk mengelompokkan entitas serupa berdasarkan fiturnya. Lalu clustering juga merupakan contoh machine learning tanpa pengawasan, dimana melatih model untuk memisahkan item ke dalam kelompok berdasarkan karakteristik, atau fiturnya semata.

Sedangkan Regression adalah Teknik supervised machine learning yang digunakan untuk memprediksi nilai numerik. Regression juga merupakan contoh teknik pada machine learning yang diawasi yang dimana kita dapat melatih model tersebut dengan menggunakan karakteristik mobil seperti ukuran mesin, jumlah kursi, jarak tempuh, dllnya untuk dapat memprediksi kemungkinan harga jualnya.

Pada model Classification yaitu bentuk machine learning yang digunakan untuk memprediksi kategori, atau kelas mana, suatu item termasuk. Seperti klinik Kesehatan yang mungkin menggunakan klasifikasi atau karakteristik pasien seperti usia, berat badan, tekanan darah, ukur tinggi, dllnya, hal ini dilakukan untuk memprediksi apakah pasien tersebut berisiko terkena diabetes.

2. Buatlah dokumen perancangan awal untuk membuat machine learning tersebut yang akan Anda ajukan kepada atasan Anda. Berisi spek mesin learning yang sesuai, bagaimana proses langganannya dan berapa harganya.

Jawaban:

Bekasi, 2 September 2022

No : 422/04/S.PNWR/2022
Lampiran : -
perihal : Pembuatan machine learning

Kepada Yth
Kepala kantor PT Telekomunikasi
Di tempat

Assalamualaikum Wr. Wb.

Salam sejahtera kami berikan kepada Bapak/Ib Pimpinan, semoga dalam menjalankan tugas dan aktivitas sehari-hari senantiasa diberikan kesehatan dan kekuatan oleh Allah SWT.

Sehubung dengan konsultasi yang telah dilakukan sebelumnya, dan setelah dipelajari lebih lanjut terkait kebutuhan yang diperlukan dalam proses pembuatan machine learning ini. Untuk membuat sebuah layanan machine learning yang dapat meneliti dan juga mengelompokkan komodo-komodo berdasarkan dengan kemiripan dan proporsi tubuh para Komodo tersebut. Dengan menggunakan machine learning model clustering yang dimana model clustering merupakan penganalsaan data yang sering dimasukkan sebagai salah satu metode data mining. Tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama.

Proses langganan ini dapat kita lihat pada service hub yang ada di dalam Microsoft Azure. Dengan langkah-langkah yang pertama klik Kesehatan lalu pilih penilaian. Kedua, klik pada tombol Pra-konfigurasi penilaian lalu klik dapatkan akun Azure. Ini

akan membuat email intruksi untuk dikirim dan untuk memulai proses pembuatan azure yang disponsori dengan Azure.

Maka dari itu kami ingin mengajukan pembuatan machine learning untuk mengelompokkan komodo-komodo tersebut berdasarkan kemiripan dan proporsi tubuh mereka dengan total pembuatan sebesar Rp. 26.716.482 terbilang (Dua Puluh Enam Juta Tujuh Ratus Enam Belas Ribu Empat Ratus Delapan Puluh Dua Ribu Rupiah).

Dengan rincian sebagai berikut:

- a. Azure SQL database
- b. Azure Machine Learning
- c. Bandwidth
- d. Storage account
- e. Microsoft Cost Management

Dengan spesifikasi machine learning:

- a. Menggunakan Model clustering
- b. Kapasitas storage 1 TB
- c. Bandwidth 5 GB dengan server US

Biaya tersebut merupakan kalkulasi penggunaan selama satu tahun penuh, dengan perhitungan Rp. 26.716.482. Perhitungan ini belum termasuk biaya jasa dan perawatan dari machine learning.

Demikian dokumen pengajuan perancangan awal untuk membuat machine learning yang akan kami jalankan, atas perhatiannya kami sampaikan terima kasih.

Wassalamualaikum Wr. Wb.

Hormat kami,
Ketua Divisi Bagian IT

Amelia Angraini M

3. Setelah atasan Anda menyetujui rancangan Anda, buatlah dokumen tentang tahapan pembuatan Machine Learning tersebut, mulai dari:

- a) Persiapan data, Bagaimana cara mendapatkan dataset dan mengidentifikasi fitur dan label dalam kumpulan data tersebut. Pra-proses, atau clean dan transform data sesuai kebutuhan.
- b) Melatih Model, Bagaimana cara membuat sebuah model dan melatihnya dengan dataset yang sudah Anda miliki.
- c) Evaluasi Performance, Bagaimana cara mengevaluasi kinerja dari Machine Learning yang sudah Anda buat.
- d) Deploy Predictive service, Bagaimana menerapkan layanan Machine learning yang sudah Anda buat menjadi sebuah pipeline yang siap digunakan untuk memprediksi data-data selanjutnya

Jawaban:

DOKUMEN TAHAPAN PEMBUATAN MACHINE LEARNING

Pada dokumen ini kami akan sajikan dokumen mengenai tahapan pembuatan machine learning dengan model clustering.

1. Persiapan Data

Di azure machine learning, data untuk pelatihan model dan operasi lainnya biasanya dienkapsulasi dalam objek yang disebut himpunan data. Kami akan menggunakan himpunan data yang mencakup pengamatan 2 spesies Komodo dengan melalui proses observasi yang sudah dilakukan, dan subset data tersebut akan disediakan dalam bentuk url web.

- a. Buat himpunan data dari file web, menggunakan pengaturan berikut:

Info dasar:

- **URL Web:** url web yang telah dibuat
- **Nama:** Komodo-data
- **Jenis himpunan data:** Tabular
- **Deskripsi:** Data Komodo
- **Lewati validasi data:** *jangan pilih*
- **Pengaturan dan pratinjau:**
- **Format file:** Dibatasi
- **Pemisah:** Koma
- **Pengodean:** UTF-8
- **Header kolom:** Hanya file pertama yang memiliki header

- **Lewati baris:** Tidak ada
- **Himpunan data berisi data multi-baris:** *jangan pilih*

Skema:

- Sertakan semua kolom selain Jalur
- Meninjau jenis yang terdeteksi secara otomatis

Konfirmasikan detail:

- Jangan membuat profil himpunan data setelah pembuatan
- Setelah himpunan data dibuat, buka dan tampilkan halaman Jelajahi untuk melihat sampel data. Data ini menunjukkan kemiripan dan proporsi tubuh untuk beberapa pengamatan komodo. Spesies komodo yang diwakili dalam himpunan data: *Varanus Komodoensis*.

b. Membuat Alur

Untuk memulai dengan perancang pembelajaran mesin Azure, pertama-tama kami harus membuat alur dan menambahkan himpunan data yang ingin kami gunakan.

c. Membuat Data Kanvas

Pada proses ini memuat skema profil data yang menyatakan bahwa kami dapat melihat distribusi berbagai kolom sebagai histogram melalui visualisasi hasil komodo data.

d. Menerapkan Transformasi

Untuk mengelompokkan pengamatan komodo, kita hanya akan menggunakan pengukuran, jadi kita akan membuang kolom spesies. Kita juga perlu menghapus baris tempat nilai hilang dan menormalkan nilai pengukuran numerik sehingga berada pada skala yang sama.

e. Menjalankan Alur Sebagai Eksperimen

2. Untuk Melatih Model

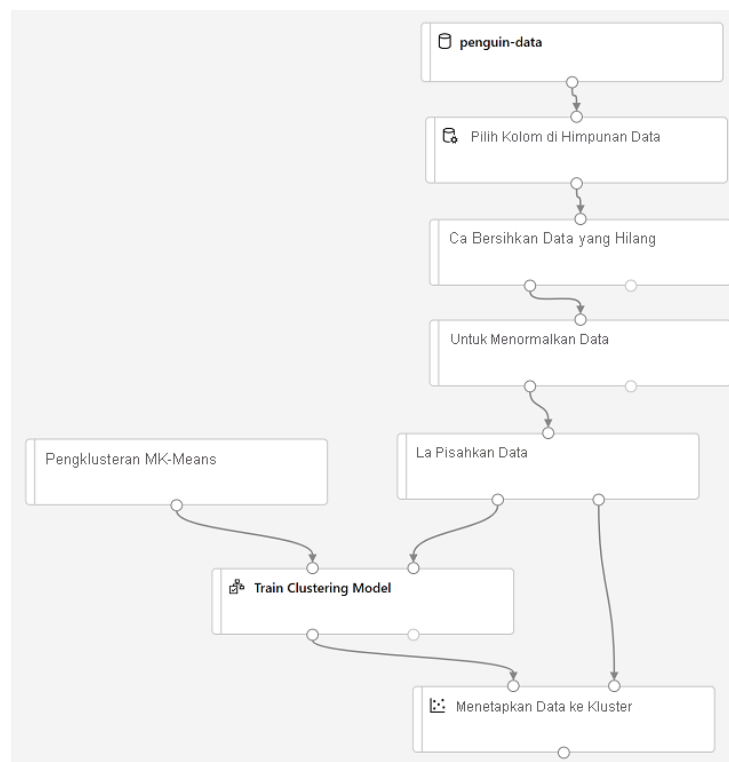
a. Menambahkan Modul Pelatihan

Untuk melatih model pengklusteran, kami perlu menerapkan algoritme pengklusteran ke data, hanya menggunakan fitur yang telah kami pilih untuk pengklusteran. kami akan melatih model dengan subset data, dan menggunakan sisanya untuk menguji model terlatih

Model yang kami latih akan menggunakan fitur untuk mengelompokkan data ke dalam kluster, jadi kami perlu melatih model menggunakan algoritme *pengklusteran*. Dalam Pustaka aset, cari dan tempatkan modul Pengklusteran K-Means ke kanvas, ke sebelah kiri himpunan data Komodo data dan di atas modul Melatih Model Pengklusteran. Kemudian, sambungkan output-nya ke input Model tidak terlatih (kiri) modul Latih Model Pengklusteran. Algoritme *K-Means* mengelompokkan item ke jumlah kluster yang Anda tentukan - nilai yang dirujuk sebagai *K*. Pilih modul Pengklusteran K-Means dan di kanan panel, atur parameter Jumlah sentroid menjadi 3.

b. Menjalankan Alur Pelatihan

Pastikan alurnya seperti berikut:



3. Evaluasi Performance

Model pengklusteran yang sukses adalah model yang mencapai tingkat pemisahan yang baik antara item di setiap antara item di setiap kluster, jadi kami membutuhkan metrik untuk membantu kami mengukur pemisahan tersebut.

Di pustaka aset, cari dan tempatkan modul mengevaluasi model pada kanvas, di bawah modul menetapkan data ke kluster. Sambungkan output

modul menetapkan data ke kluster ke input himpunan data yang dinilai (kiri) dari modul mengevaluasi model.

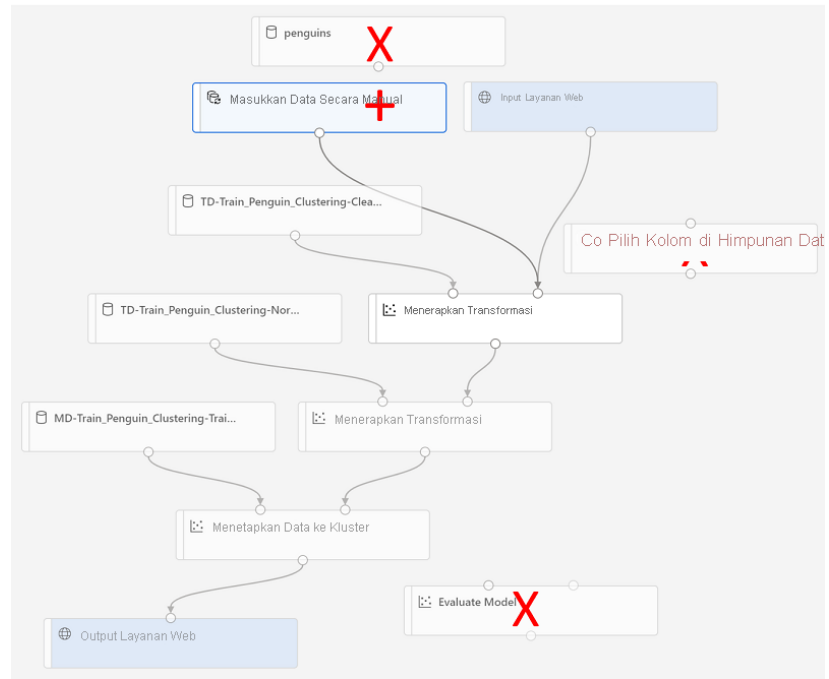
Matriks di setiap baris adalah:

- **Jarak Rata-Rata ke Pusat Lain:** Ini menunjukkan seberapa dekat, rata-rata, setiap titik dalam kluster dengan sentroid dari semua kluster lainnya.
- **Jarak Rata-rata ke Pusat Kluster:** Ini menunjukkan seberapa dekat, rata-rata, setiap titik dalam kluster dengan sentroid kluster.
- **Jumlah Titik:** Jumlah titik yang ditetapkan ke kluster.
- **Jarak Maksimal ke Pusat Kluster:** Maksimum jarak antara setiap titik dan sentroid dari kluster titik itu. Jika jumlah ini tinggi, kluster mungkin tersebar luas. Statistik ini yang dikombinasikan dengan Jarak Rata-Rata ke Pusat Kluster membantu kami menentukan *penyebaran* kluster.

4. Deploy Predictive Service

- Ganti himpunan data Komodo dengan modul masukkan data secara manual yang tidak menyertakan kolom spesies.
- Hapus modul lalu pilih kolom dalam himpunan data, yang sekarang redundan.
- Sambungkan modul input layanan web dan masukkan data secara manual (yang mewakili input data yang akan diklusterkan) ke modul terapkan transformasi pertama.
- Hapus modul evaluasi model.

Contoh Alur:



Setelah alur selesai, pilih detail pekerjaan, di jendela baru klik kanan modul menetapkan data ke kluster dan pilih lakukan pratinjau data untuk melihat penetapan dan matriks kluster yang diprediksi untuk tiga pengamatan komodo dalam data input.

Alur inferensi kami menetapkan pengamatan komodo ke kluster berdasarkan fiturnya. Sekarang kami siap untuk menerbitkan alur sehingga aplikasi klien dapat menggunakannya.