

LAPORAN PRAKTIKUM BIG DATA ANALYTICS
FAKTOR PENENTU KESEGARAN IKAN
STUDI KORELASI ANTARA KONDISI PERAIRAN, METODE
PENANGKAPAN, DAN KARAKTERISTIK IKAN

(Laporan Ini Disusun Guna Memenuhi Tugas Akhir Praktikum Big Data Analytics)

Dosen Pengampu: Agus Nugraha, S.T., M.Kom.

Asisten Laboratorium: Ryan Hakim & Silvi Nurinsan



Disusun Oleh:

| | |
|-----------------------|-----------|
| Fitri Fatimah | 2C2230004 |
| Nadine Valia Azzahra | 2C2230007 |
| Kanaya Dzikra Setiadi | 2C2230015 |
| Amelia Regina Permana | 2C2230019 |

PROGRAM STUDI SAINS DATA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS KOPERASI INDONESIA

2025

1. Latar Belakang

Kualitas hasil tangkapan laut, khususnya tingkat kesegaran ikan, merupakan salah satu faktor penting yang menentukan nilai jual, kelayakan konsumsi, dan daya saing produk perikanan di pasar lokal maupun global. Kesegaran ikan tidak hanya mencerminkan mutu produk, tetapi juga menjadi tolok ukur efisiensi dan higienisasi proses penangkapan serta penanganan pasca-tangkap. Dalam praktiknya, tingkat kesegaran sangat dipengaruhi oleh kombinasi faktor lingkungan seperti suhu perairan, salinitas, dan kedalaman, serta metode penangkapan yang diterapkan. Selain itu, karakteristik biologis dari masing-masing jenis ikan juga berperan penting dalam menentukan kecepatan penurunan kualitas setelah ditangkap. Di tengah meningkatnya permintaan terhadap produk perikanan yang segar dan berkualitas tinggi, penting untuk memahami secara ilmiah faktor-faktor apa saja yang paling berkontribusi terhadap kesegaran ikan, agar praktik perikanan dapat dioptimalkan secara berkelanjutan. Pemahaman ini juga mendukung upaya pengurangan pemborosan hasil laut dan meningkatkan efisiensi rantai distribusi. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengkaji secara menyeluruh hubungan antara kesegaran ikan dan berbagai faktor yang memengaruhinya.

2. Tujuan Penelitian

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis hubungan antara tingkat kesegaran ikan dengan berbagai faktor yang memengaruhinya, khususnya kondisi lingkungan perairan (seperti suhu, salinitas, dan kedalaman laut), metode penangkapan yang digunakan, serta karakteristik ikan berdasarkan jenis dan kategorinya. Dengan pendekatan korelasi data, penelitian ini diharapkan dapat memberikan gambaran menyeluruh mengenai faktor-faktor dominan yang memengaruhi kesegaran ikan, sekaligus menjadi dasar pertimbangan dalam pengembangan kebijakan atau strategi peningkatan mutu hasil tangkapan laut. Selain itu, hasil dari penelitian ini diharapkan dapat membantu nelayan, pengolah ikan, serta pemangku kepentingan lainnya dalam mengambil keputusan berbasis bukti. Dengan begitu, keberlanjutan ekosistem laut dan kualitas produk perikanan dapat lebih terjaga dari hulu hingga ke konsumen.

3. Eksplorasi Dataset Ikan

Tahapan awal dalam penelitian ini adalah melakukan eksplorasi terhadap dataset ikan yang menjadi fokus utama analisis. Dataset ini merupakan data hasil tangkapan ikan dari berbagai kategori, metode penangkapan, serta lokasi pasar. Beberapa kolom utama yang terdapat dalam dataset ini meliputi `fish_name` (nama ikan), `category_name` (kategori ikan seperti pelagis atau demersal), `weight_kg` (berat dalam kilogram), `avg_price_per_kg` (harga rata-rata per kilogram), `fishing_method` (metode penangkapan yang digunakan), `fisher_name` (nama nelayan), `boat_type` (jenis kapal), dan `market_location` (lokasi pasar ikan).

Langkah pertama dalam eksplorasi adalah menampilkan dimensi dataset, yaitu jumlah baris dan kolom, untuk memahami seberapa besar data yang akan dianalisis. Kemudian dilakukan pengecekan terhadap jenis data pada setiap kolom untuk memastikan kesesuaian tipe data terhadap operasi yang akan dilakukan. Misalnya, kolom harga dan berat harus bertipe numerik agar dapat dihitung rata-ratanya.

Selanjutnya, dilakukan identifikasi terhadap nilai kosong (missing values), data duplikat, dan nilai ekstrem yang mungkin tidak masuk akal (seperti berat ikan 0 kg atau harga negatif). Visualisasi awal seperti histogram atau countplot digunakan untuk melihat distribusi nilai pada kolom kategorikal dan numerik. Proses eksplorasi ini penting karena memberikan pemahaman awal terhadap kualitas data serta membantu menyusun strategi pra-pemrosesan dan analisis lanjutan yang lebih tepat sasaran.

4. Premrosesan dan Analisis Data Menggunakan Pandas dan PySpark

Pada tahap ini, dilakukan proses penggabungan data (join) untuk mengintegrasikan informasi dari dua DataFrame berbeda, yaitu `fish_dataset` dan `fish_categories_dataset`. Tujuan dari penggabungan ini adalah untuk memperoleh informasi yang lebih lengkap terkait dengan masing-masing ikan berdasarkan kategorinya.

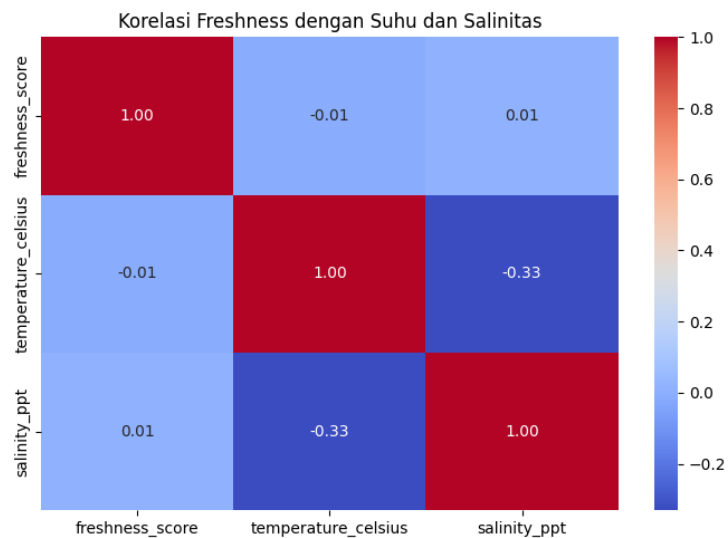
4.1 Join Operations

Untuk menggabungkan kedua DataFrame tersebut, digunakan metode `.join()` dari pustaka Pandas dengan parameter `on="category_id"` dan `how="inner"`. Operasi Inner Join dipilih karena hanya akan mengambil data yang memiliki nilai `category_id` yang sama di kedua DataFrame. Artinya, hanya data ikan yang memiliki kecocokan kategori dengan dataset kategori ikan yang akan dimasukkan dalam hasil akhir.

Hasil dari proses ini disimpan dalam sebuah DataFrame baru bernama `joined_df`, yang kini memuat informasi ikan sekaligus kategori dan deskripsi dari masing-masing kategori. Dengan data yang telah digabungkan, proses analisis selanjutnya dapat dilakukan dengan lebih efektif, karena seluruh informasi relevan telah tersedia dalam satu struktur data yang utuh untuk data berbentuk file yang sudah di *joined* *terlampir*.

4.2 Visualisasi Data Hasil Pemrosesan

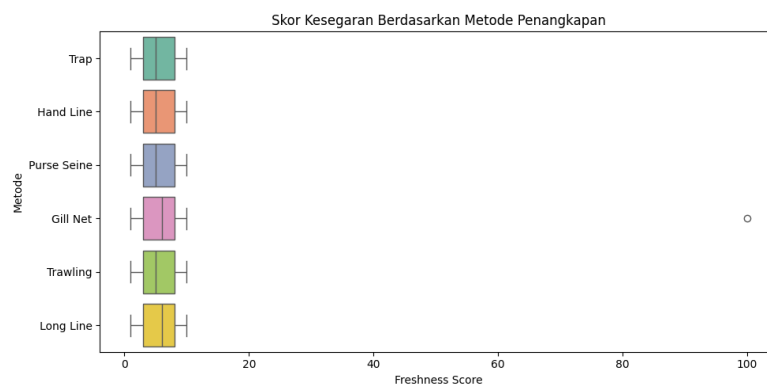
A. HEATMAP KORELASI KONDISI PERAIRAN & KESEGARAN IKAN



Gambar 1. HeatMap Kolerasi Kondisi Perairan & Kesegaran Ikan.

Gambar 1. Berdasarkan heatmap korelasi, dapat dilihat bahwa suhu air laut dan salinitas tidak memiliki hubungan yang signifikan dengan freshness score ikan, dengan nilai korelasi masing-masing hanya -0,01 dan 0,01. Artinya, tingkat kesegaran ikan tidak dipengaruhi secara langsung oleh suhu maupun kadar garam air laut. Sementara itu, korelasi antara suhu dan salinitas sebesar -0,33 menunjukkan hubungan negatif sedang, di mana saat suhu naik, salinitas cenderung menurun. Secara keseluruhan, kesegaran ikan lebih dipengaruhi oleh faktor lain seperti metode penanganan pasca-penangkapan.

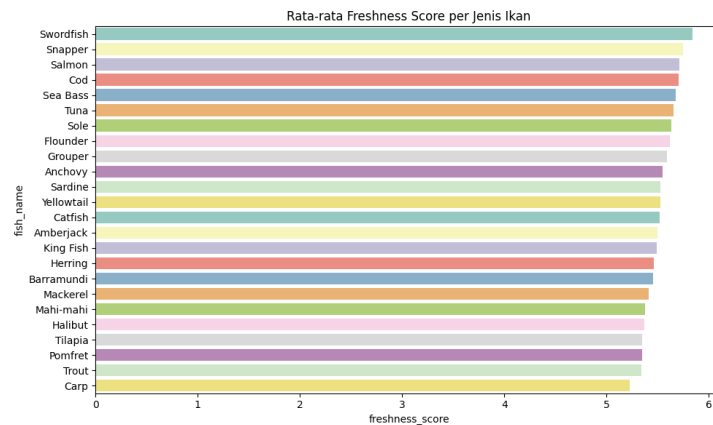
B. BOXPLOT KESEGARAN IKAN PER METODE PENANGKAPAN



Gambar 2. BoxPlot Kesegaran Ikan Per Metode Penangkapan.

Boxplot menunjukkan bahwa semua metode penangkapan menghasilkan skor kesegaran yang mirip, dengan median rendah di kisaran 5–10. Terdapat satu outlier ekstrem pada metode Gill Net dengan skor hampir 100, namun itu tidak mewakili keseluruhan. Artinya, metode penangkapan tidak berpengaruh signifikan terhadap kesegaran ikan, dan faktor lain seperti penyimpanan atau waktu penanganan lebih berperan penting.

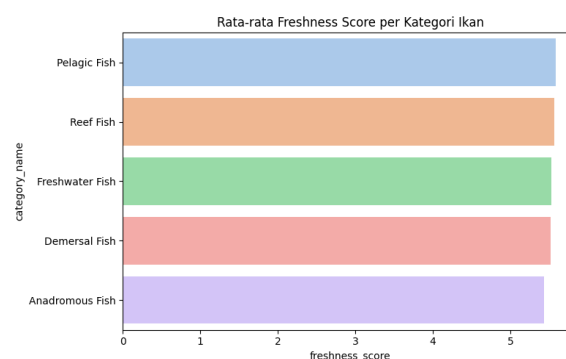
C. BARPLOT RATA-RATA KESEGARAN IKAN PER JENIS IKAN



Gambar 3. BarPlot Rata-Rata Kesegaran Ikan Per Jenis Ikan.

Grafik menunjukkan bahwa Swordfish, Snapper, dan Salmon memiliki rata-rata skor kesegaran tertinggi dibandingkan jenis ikan lainnya, yang mengindikasikan bahwa ketiganya kemungkinan ditangani dengan lebih baik atau lebih cepat setelah penangkapan. Di sisi lain, ikan seperti Carp, Trout, dan Tilapia memiliki skor kesegaran yang lebih rendah, meskipun selisihnya tidak terlalu besar. Secara umum, grafik ini mengungkapkan bahwa jenis ikan dapat memengaruhi tingkat kesegaran, kemungkinan dipengaruhi oleh faktor seperti ketahanan ikan, habitat, dan proses distribusinya.

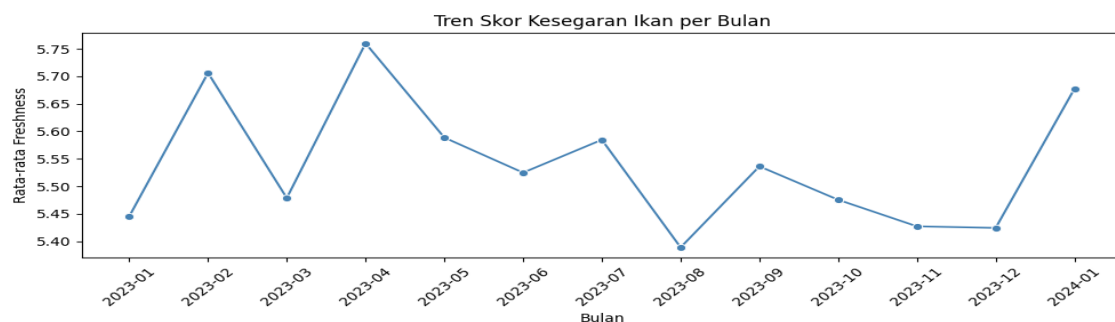
D. BARPLOT KESEGARAN IKAN PER KATEGORI IKAN



Gambar 4. BarPlot Kesegaran Ikan Per Kategori ikan.

Grafik menunjukkan rata-rata skor kesegaran berdasarkan kategori ikan. Pelagic Fish memiliki rata-rata skor tertinggi, diikuti oleh Reef Fish, Freshwater Fish, dan Demersal Fish, sementara Anadromous Fish memiliki skor terendah. Hal ini mengindikasikan bahwa ikan yang berasal dari perairan terbuka cenderung lebih segar saat ditangkap. Perbedaan ini kemungkinan dipengaruhi oleh lokasi habitat, durasi transportasi, dan metode penanganan pasca-penangkapan. Dengan demikian, kategori ikan turut berperan dalam menentukan tingkat kesegarannya.

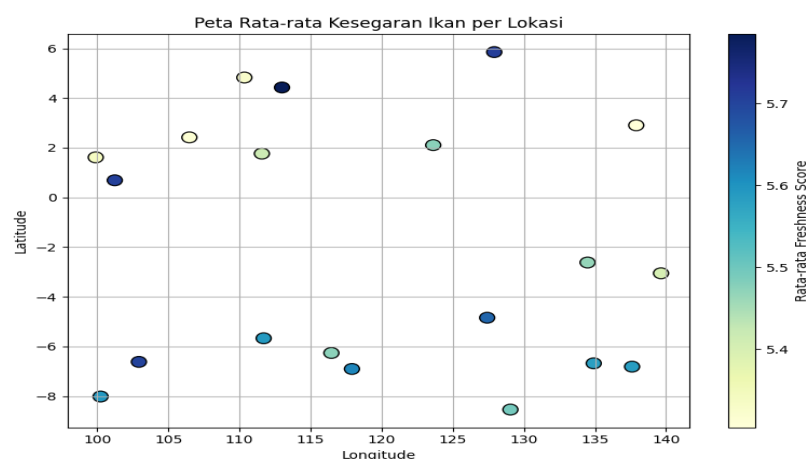
E. LINE CHART TREN KESEGARAN IKAN PER BULAN



Gambar 5. Tren Skor Kesegaran Ikan per Bulan

Berdasarkan grafik, rata-rata skor kesegaran ikan berfluktuasi sepanjang tahun 2023. Nilai tertinggi tercatat pada April 2023, sedangkan titik terendah terjadi pada Agustus 2023. Penurunan kualitas pada bulan-bulan tertentu kemungkinan dipengaruhi oleh kondisi lingkungan atau distribusi. Sementara itu, terjadi kenaikan signifikan di Januari 2024, yang dapat menandakan awal musim tangkap baru atau perbaikan sistem distribusi. Pola ini penting untuk mengevaluasi waktu panen dan pengelolaan rantai pasok ikan secara optimal.

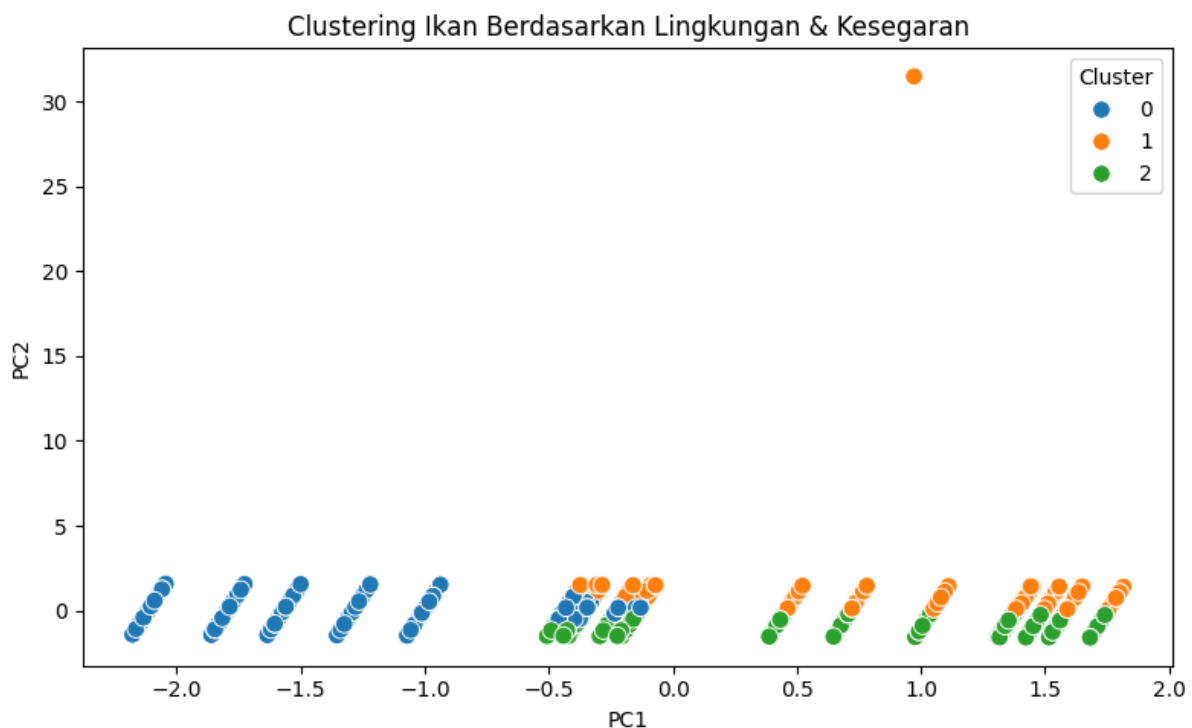
F. PETA RATA-RATA KESEGARAN IKAN PER LOKASI



Gambar 6. Peta Rata-rata Kesegaran Ikan per Lokasi.

Peta di atas menunjukkan variasi skor kesegaran ikan berdasarkan lokasi geografis. Warna biru tua menandakan nilai kesegaran yang lebih tinggi, sementara warna kekuningan menunjukkan nilai yang lebih rendah. Terlihat bahwa skor kesegaran cenderung lebih tinggi di wilayah utara dan selatan, sedangkan wilayah tengah memiliki skor yang lebih rendah. Perbedaan ini kemungkinan dipengaruhi oleh kondisi lingkungan seperti suhu dan salinitas, serta faktor distribusi seperti jarak tempuh dan kualitas penanganan pasca-tangkap. Informasi ini dapat dimanfaatkan untuk mengidentifikasi daerah yang membutuhkan perbaikan dalam pengelolaan rantai pasok hasil perikanan.

G. CLUSTERING IKAN BERDASARKAN LINGKUNGAN DAN KESEGARAN



Gambar 7. Clustering Ikan Berdasarkan Lingkungan & Kesegaran.

Gambar di atas menunjukkan hasil clustering ikan menggunakan dua komponen utama (PC1 dan PC2) yang diperoleh dari reduksi dimensi terhadap variabel lingkungan dan skor kesegaran. Terbentuk tiga cluster yang terpisah cukup jelas, masing-masing ditandai dengan warna berbeda.

- **Cluster 0 (biru)** cenderung memiliki nilai PC1 negatif, yang dapat diasosiasikan dengan kondisi lingkungan tertentu atau tingkat kesegaran yang lebih rendah.
- **Cluster 1 (oranye)** tersebar di tengah hingga kanan grafik, dengan satu outlier ekstrem pada PC2 tinggi. Cluster ini kemungkinan mencerminkan kelompok ikan dengan variasi kondisi lingkungan dan kesegaran yang lebih kompleks.

- **Cluster 2 (hijau)** dominan di sisi kanan, menunjukkan pola karakteristik lingkungan dan kesegaran yang berbeda dari dua cluster lainnya.

Hasil ini menunjukkan bahwa data ikan dapat dikelompokkan secara alami berdasarkan fitur lingkungan dan kesegaran, yang dapat dimanfaatkan untuk segmentasi wilayah penangkapan atau strategi distribusi yang lebih tepat sasaran.

5. Kesimpulan

Berdasarkan hasil eksplorasi dan analisis data perikanan menggunakan teknik big data analytics, dapat disimpulkan bahwa tingkat kesegaran ikan tidak secara signifikan dipengaruhi oleh kondisi perairan seperti suhu dan salinitas, karena hubungan antar variabel tersebut tergolong sangat lemah. Meskipun suhu dan salinitas memiliki korelasi negatif moderat satu sama lain, keduanya tidak memiliki hubungan yang berarti terhadap skor kesegaran. Demikian pula, metode penangkapan dan jenis kapal juga tidak menunjukkan pengaruh dominan terhadap kesegaran, meskipun beberapa jenis ikan seperti Swordfish dan Snapper cenderung memiliki nilai kesegaran yang lebih tinggi, kemungkinan besar karena perbedaan dalam penanganan pasca-tangkap. Visualisasi spasial dan temporal menunjukkan adanya fluktuasi skor kesegaran sepanjang waktu dan antar lokasi, yang mengindikasikan bahwa faktor logistik, musim tangkap, serta sistem distribusi memiliki peran penting dalam menjaga mutu ikan.

Selain itu, analisis clustering menunjukkan bahwa data ikan dapat dikelompokkan berdasarkan kombinasi variabel lingkungan dan kesegaran, yang membuka peluang untuk strategi segmentasi pasar atau distribusi berbasis karakteristik wilayah. Temuan lainnya dari proporsi berat tangkapan berdasarkan jenis kapal memperlihatkan perbedaan kontribusi signifikan antar jenis armada, memberikan informasi penting bagi evaluasi efisiensi operasional. Secara keseluruhan, hasil penelitian ini menegaskan pentingnya optimalisasi penanganan pasca-tangkap, distribusi, dan pengelolaan rantai pasok, dibandingkan hanya mengandalkan faktor lingkungan alamiah. Pendekatan berbasis data seperti ini sangat penting dalam mendukung pengambilan keputusan strategis untuk meningkatkan kualitas hasil perikanan secara berkelanjutan.

Lampiran

- Lampiran A : Program Pemrosesan Data PySpark

[Copy of Project Big Data \(pyspark\).ipynb - Colab](#)

- Lampiran B : Program Pemrosesan Data Pandas

https://colab.research.google.com/drive/17MYLO0H_V-A7TFgGd00KHxj6k0j6V8ZB?usp=sharing

- Lampiran C : Visualisasi Data dan analisis

https://colab.research.google.com/drive/160G6zslfSpr_5gzJJUi852DAYUbqnhJk?usp=sharing

https://colab.research.google.com/drive/118pHCovy_VEAnu9Cw5_ojrWRMqRFyQY_?usp=sharing

- Lampiran D : File Joined Fish Datasets

https://drive.google.com/drive/folders/1FEFd3VBWEuxDxjVeROrhQFh5UgoBj426?usp=drive_link