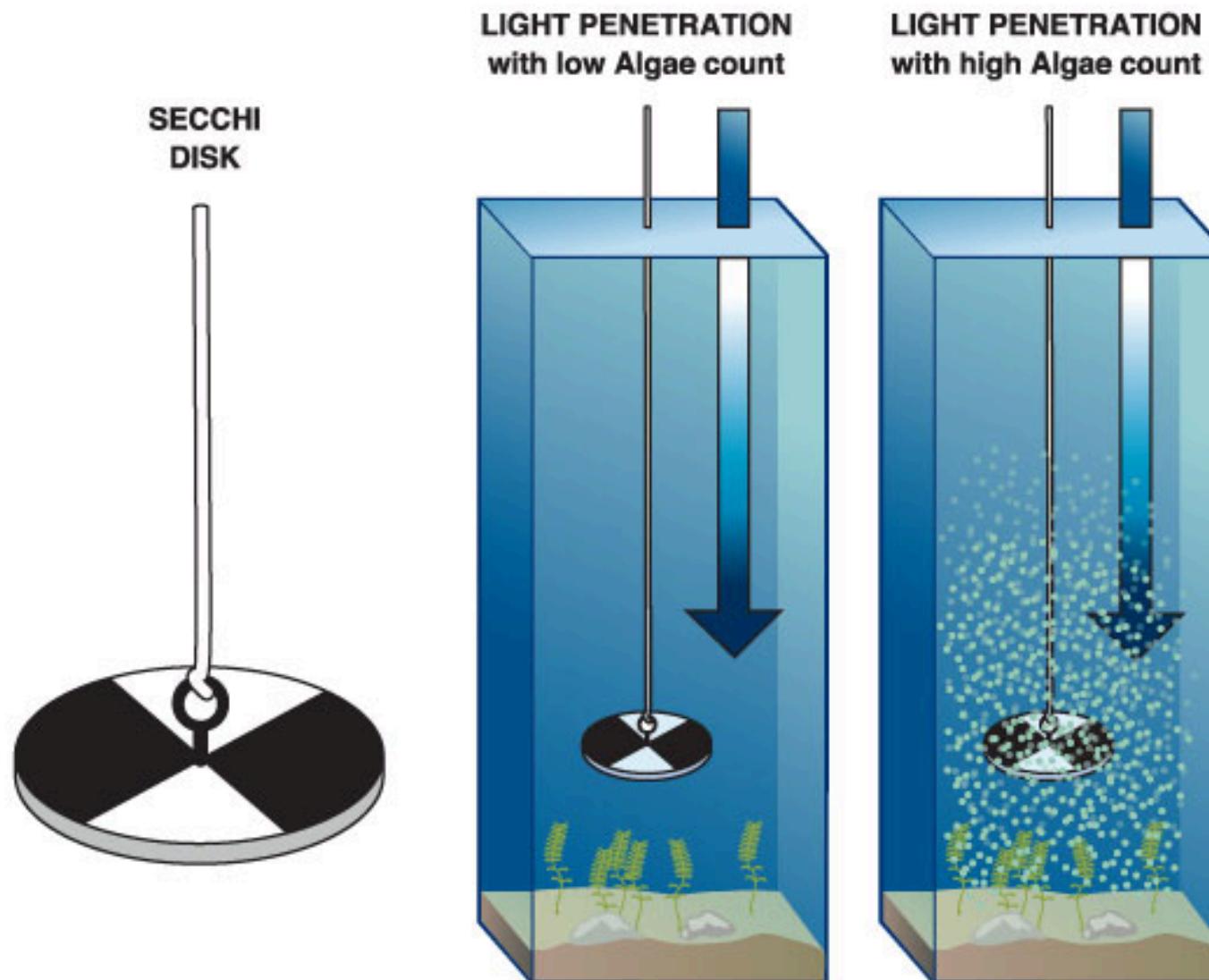


Analisis Faktor Dominan dan Pemodelan Spasial untuk Prediksi Kejernihan Air Danau

Oleh PoLaLi

Data Science Competition | Fortex 6.0

Pendahuluan



“One simple measure tells us about ecosystem health.”

Kejernihan Danau itu Penting!

- Kesehatan Ekosistem
- Rekreasi & Ekonomi
- Kebijakan & Pemantauan
- Sederhana & Konsisten

Rumusan Masalah



Dampak Ekologi dan Hayati



Pembatasan rekreasi & Wisata



Tantangan pengelolaan dan
Penganggaran



Tujuan Utama



Prediksi Berbasis Data



Prediksi Berbasis Spasial

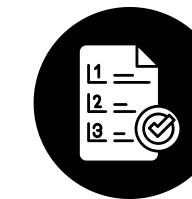


Pengklasifikasian hasil prediksi

Manfaat



Pemantauan Ekosistem Berkelanjutan



Menetapkan prioritas restorasi danau



Merencanakan strategi bisnis berbasis data

Pembahasan

Landasan Teori

- Kejernihan air danau diukur dengan Secchi depth, dipengaruhi oleh faktor fisik (kedalaman, luas), kualitas air (nutrien, sedimen, alga), serta penggunaan lahan sekitar (urban, pertanian).
- Dataset skala besar memungkinkan pemodelan prediksi kejernihan dengan teknik *machine learning*.
- Interpretabilitas model menggunakan SHAP & PDP untuk menjelaskan pengaruh tiap faktor terhadap hasil prediksi.
- Ketidakpastian prediksi ditangani dengan Conformal Prediction, sehingga setiap prediksi dilengkapi interval kepercayaan yang reliabel.

Fitur dan Target

Target**Mean Secchi****Fitur Numerik**

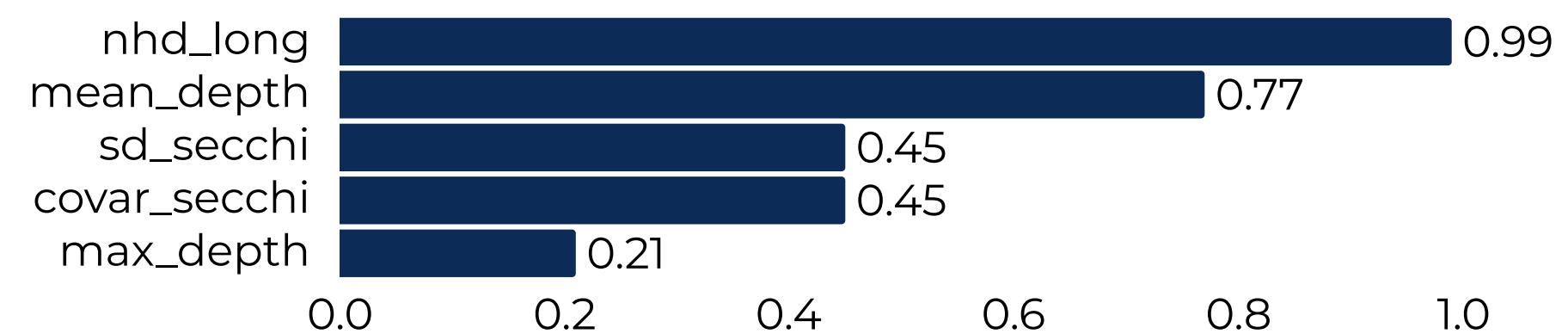
- nhd_lat
- nhd_long
- lake_area_ha
- lake_perim_meters
- meandepth
- maxdepth
- iws_area_ha
- pcturbancombined_hu12
- pctagcombined_hu12
- pctpasturehay_hu12
- pctrowcrop_hu12

Fitur Kategorik

- lagosname1
- in_nwi
- state_name
- lakeconnectivity
- glacial

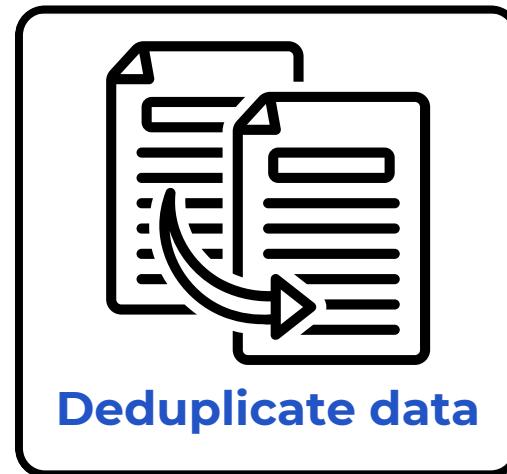
Overview Data

Dataset menggabungkan morfologi danau, karakteristik daerah aliran sungai, penggunaan lahan, dan koordinat geospasial dari ribuan danau di AS (2002–2011).

Total Baris = 4984**Total Kolom = 29****Top 5 Missing Values**

Data Cleaning

Membuang baris duplikat



Hapus pemisah ribuan (,) dan ganti (,) ke (.)



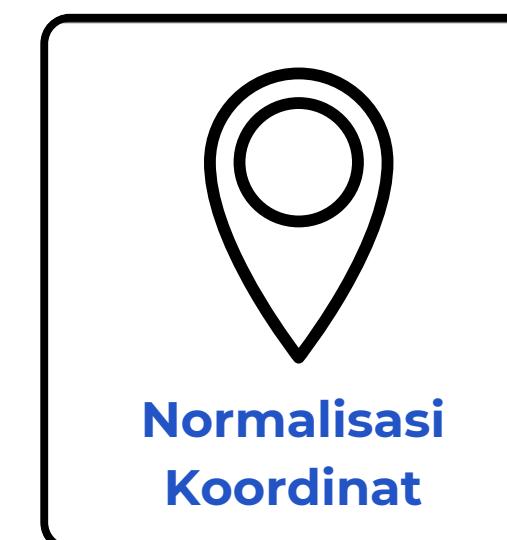
14.234.543 —————→ 14,234,543

Membuang baris dupli filter +
clip(0.1, 30.0)kat



mean_secchi > 0, cap ke [0.1, 30] m

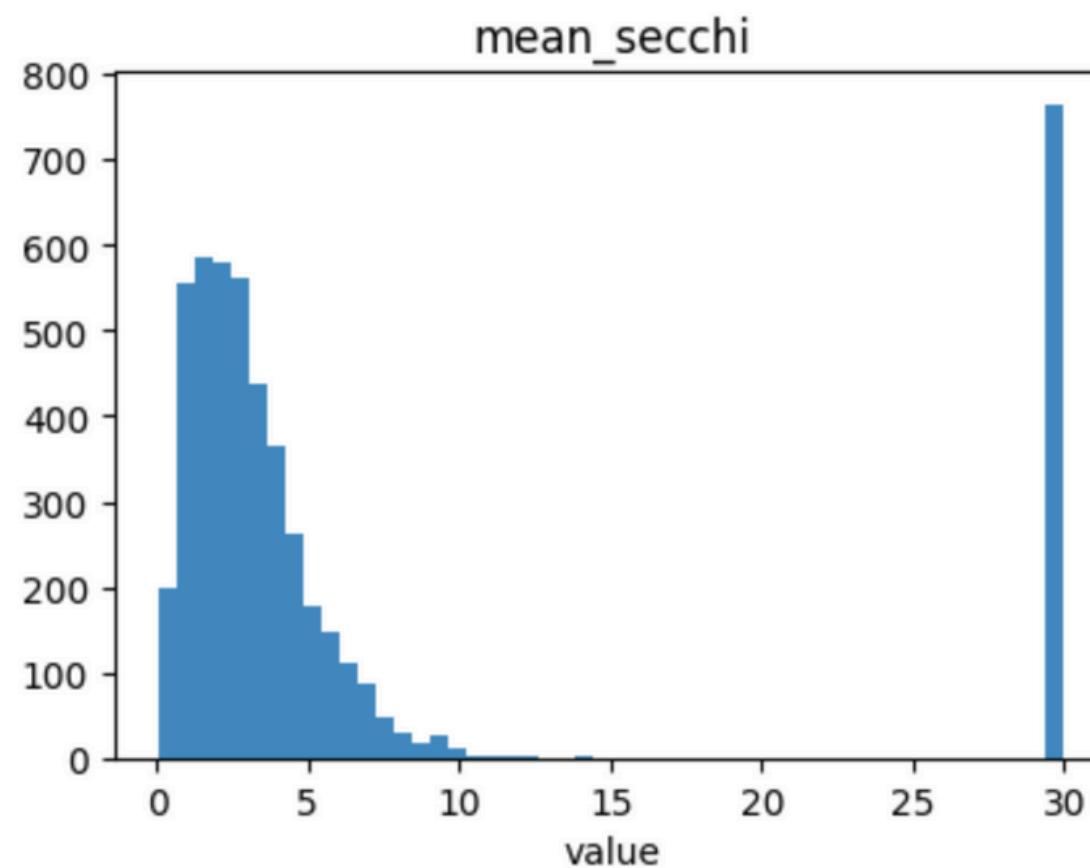
Drop data selain -90≤lat≤90,
-180≤lon≤180



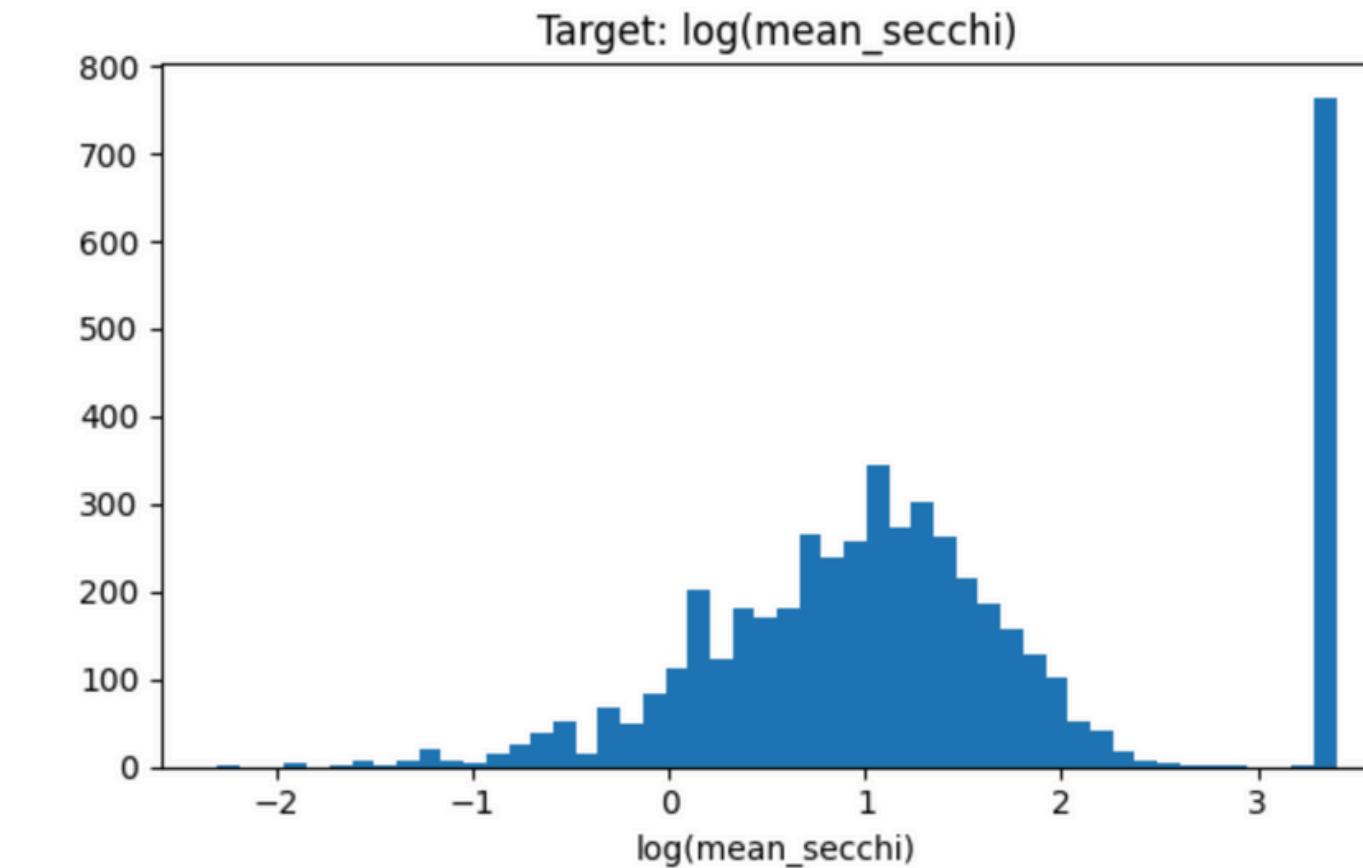
-94.660.011 —————→ -94,6

Distribusi Target & Transformasi Log

Distribusi Awal



Transformasi Log



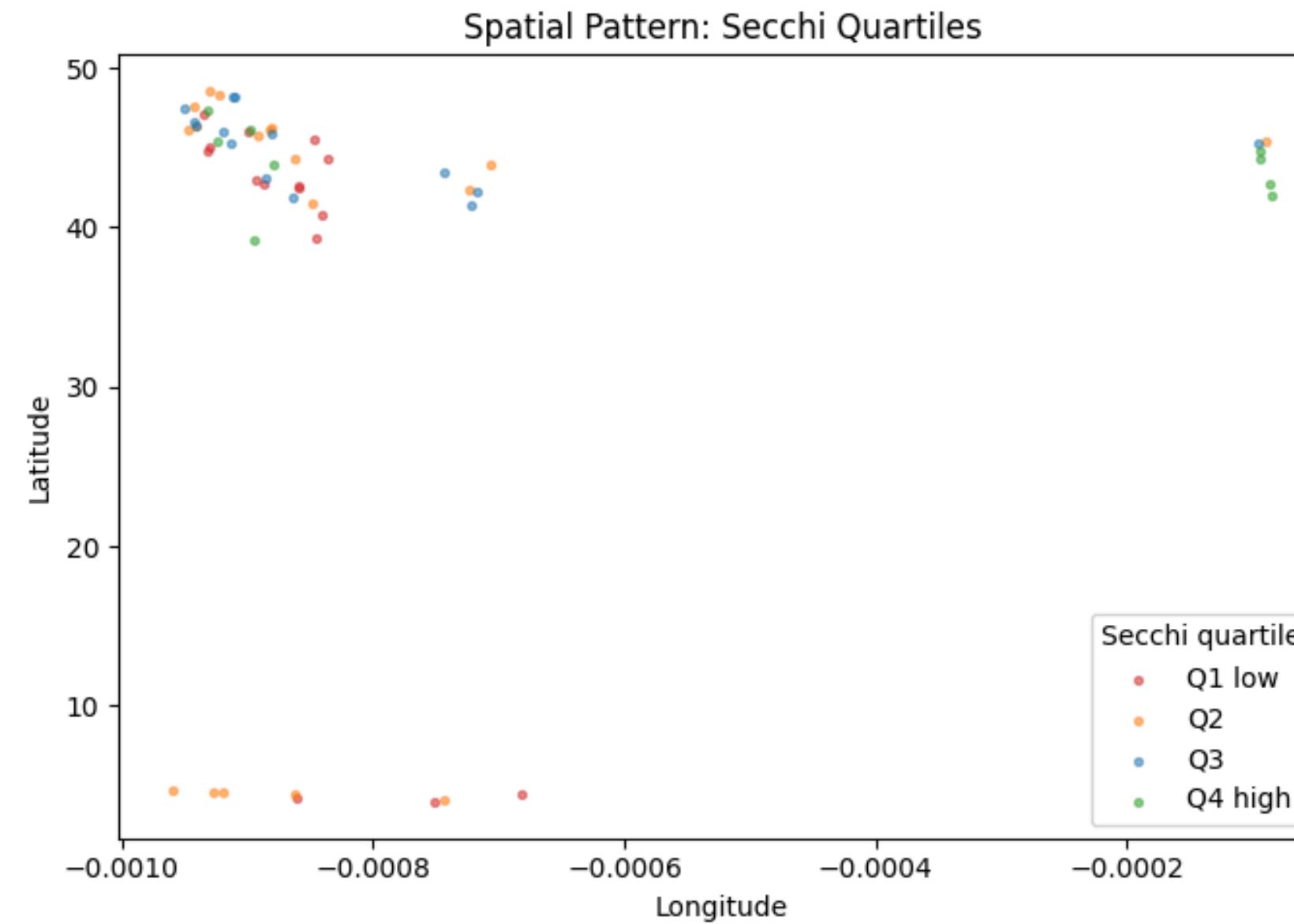
Distribusi Mean Secchi sangat miring (right skewed), transformasi log menstabilkan varians dan meningkatkan kesesuaian model.

Data Understanding

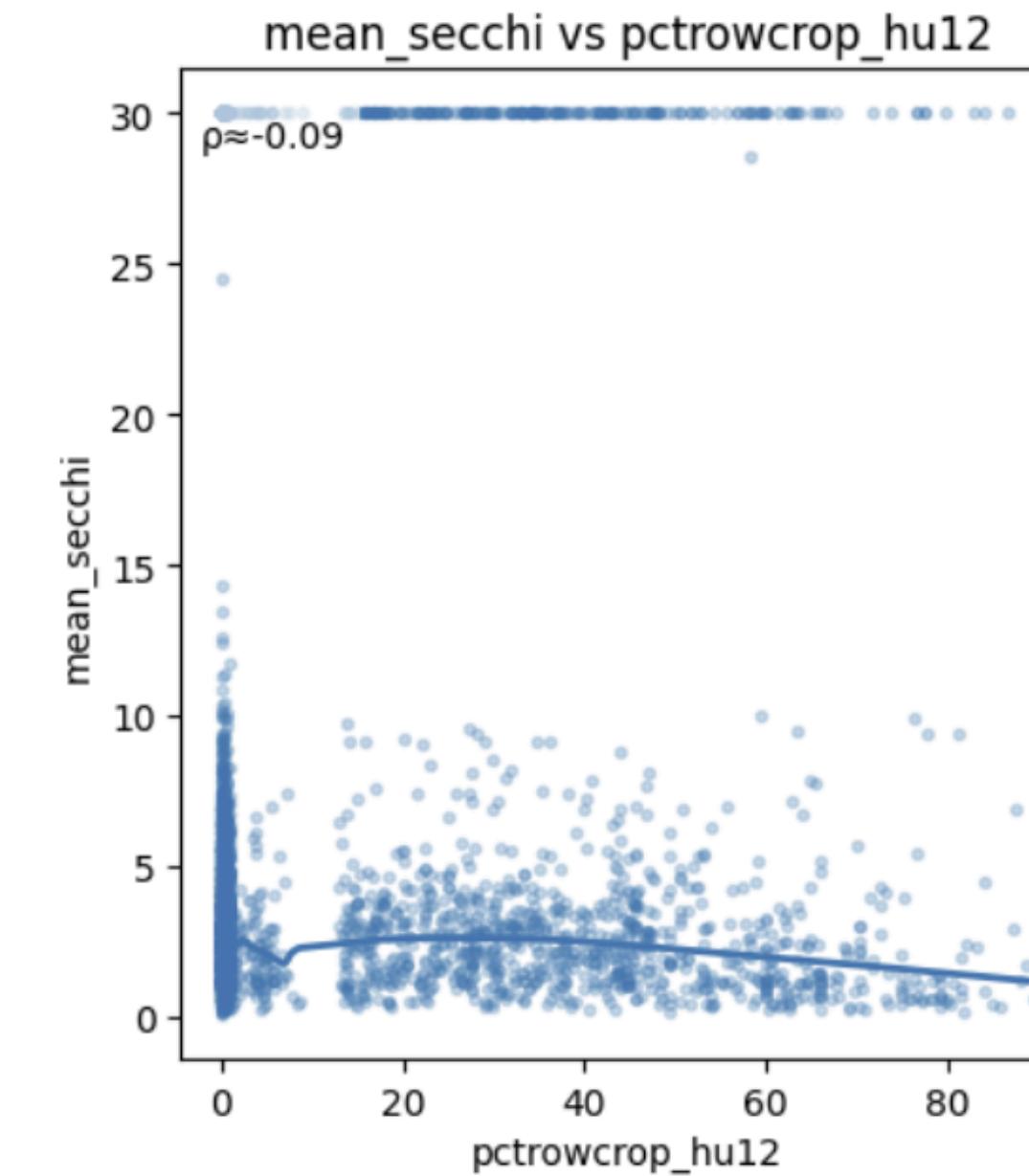
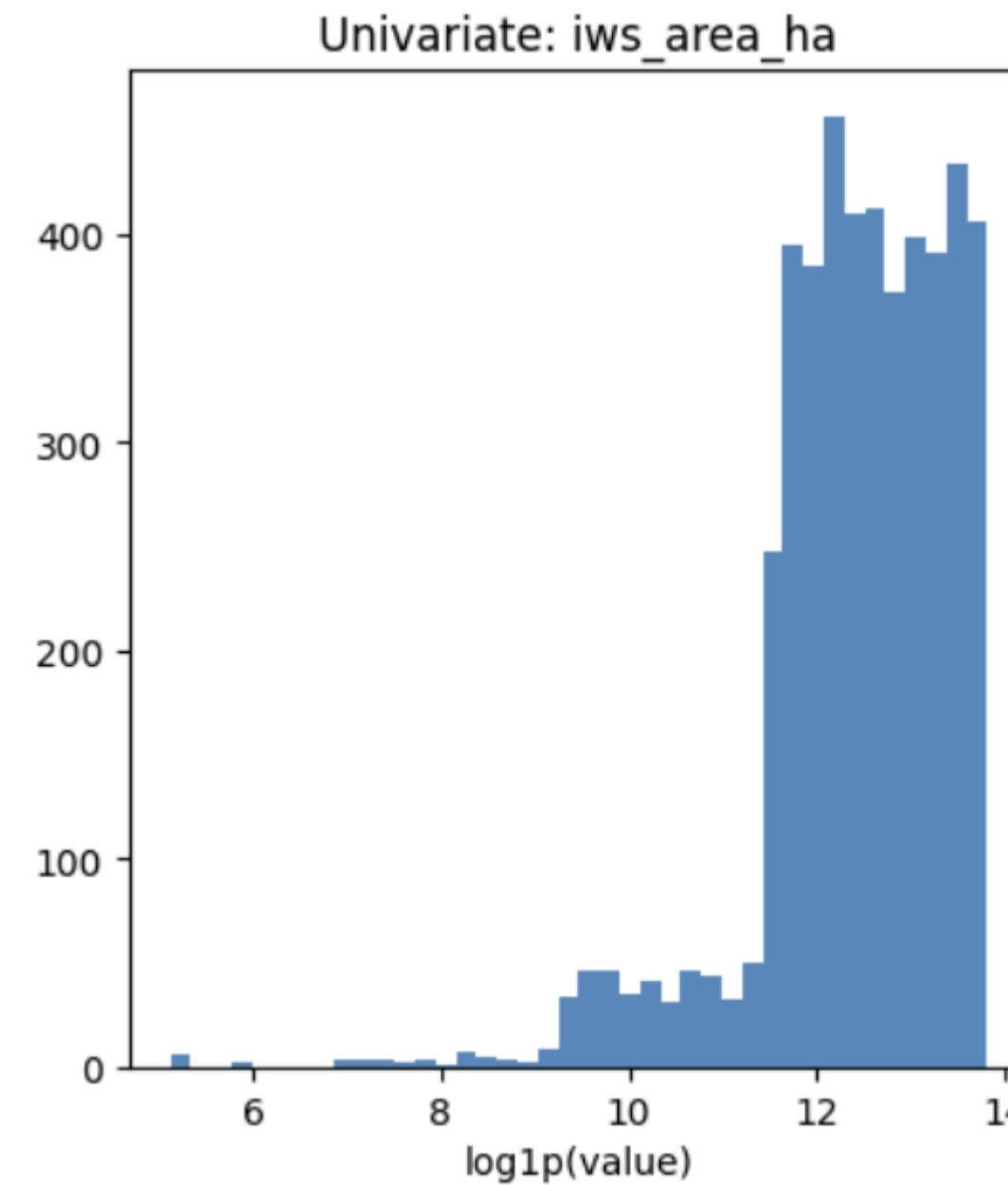
Dashboard Interaktif (Tableau)

Exploratory Insights

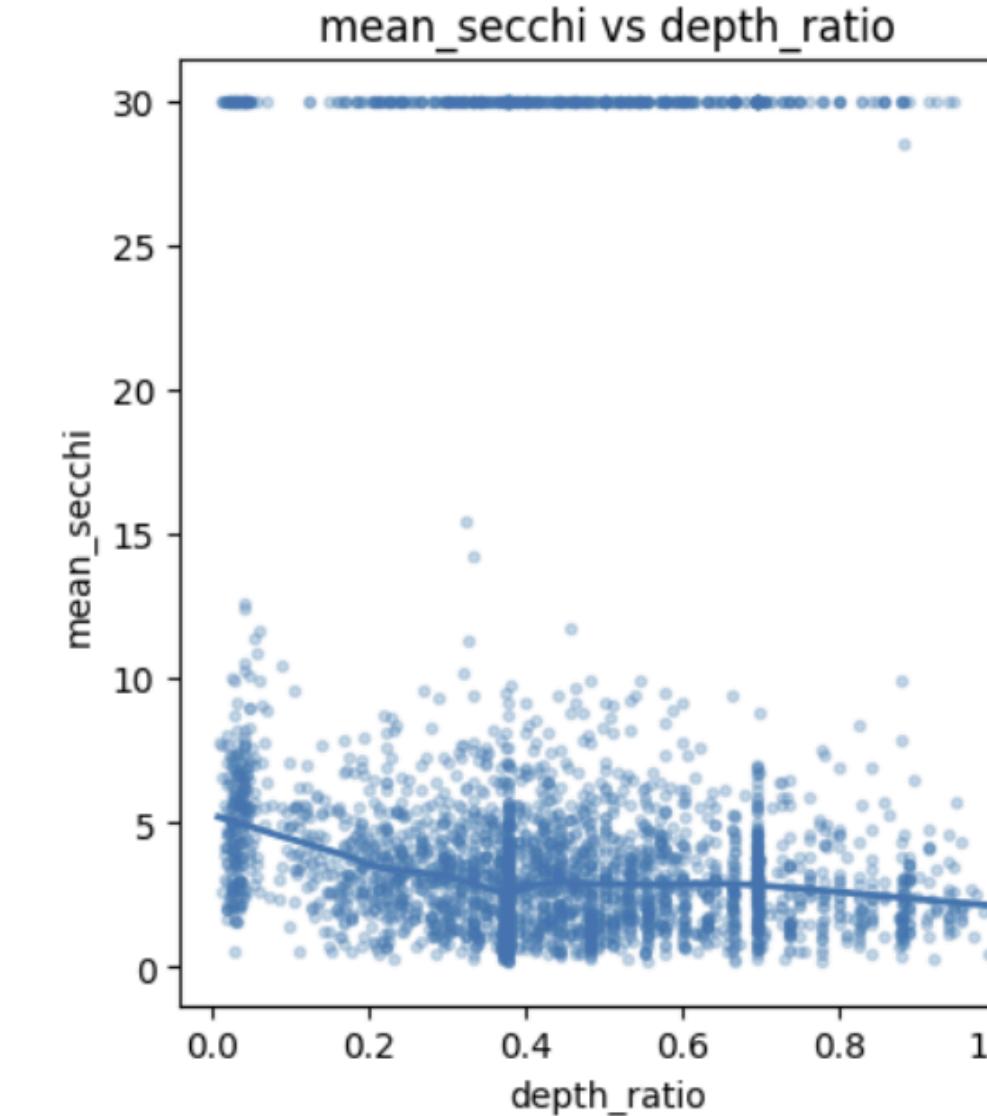
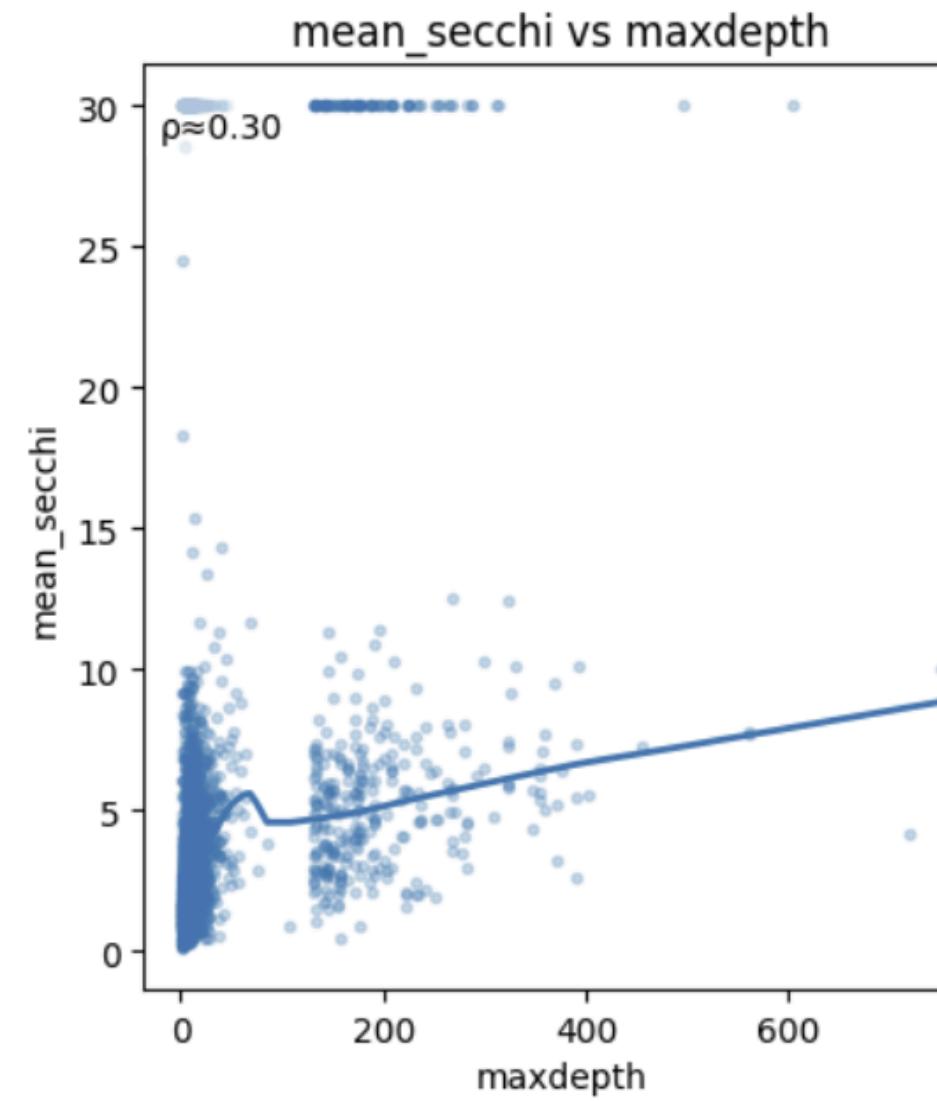
Pola Spasial



Univariate & Bivariate



Korelasi & Hubungan Awal



Semakin dalam danau, semakin baik kejernihan air

Feature Engineering

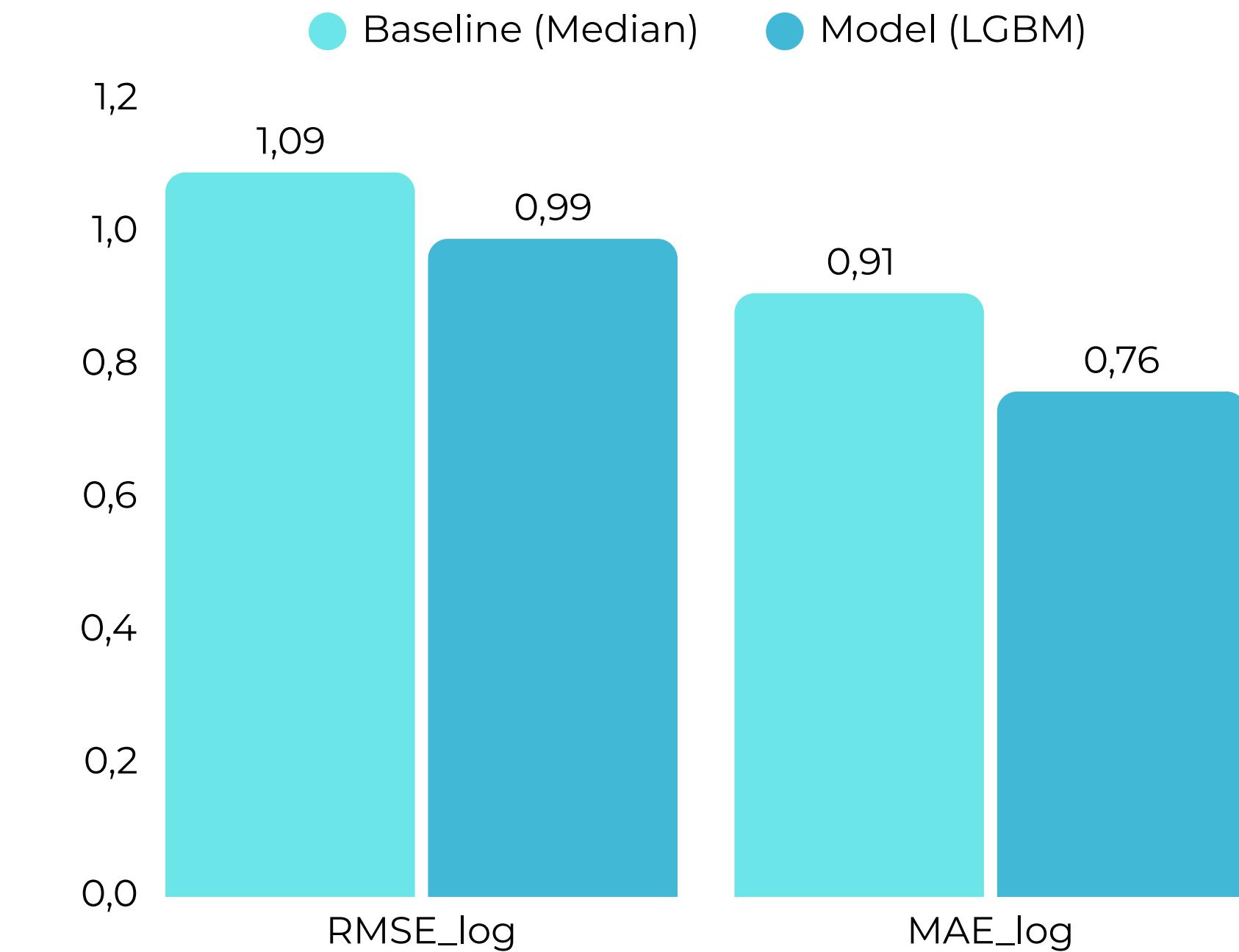
Mengubah fitur mentah menjadi representasi yang lebih kaya, merefleksikan proses hidrologis dan geografis yang memengaruhi kejernihan air danau

Raw Features	Engineering	Engineered Features
<ul style="list-style-type: none">• Morfologi Danau• Penggunaan Lahan• Daerah Aliran Sungai (DAS)• Koordinat	<ul style="list-style-type: none">• Kombinasi & Rasio Fitur• <i>Domain-Specific Feature Engineering</i>• Transformasi Trigonometri• Transformasi Logaritmik	<ul style="list-style-type: none">• depth_ratio• SDI (<i>Shoreline Development Index</i>)• ag_intensity• Fitur Spasial (sin cos)

Proses FE meningkatkan kualitas informasi fitur sehingga model lebih mampu menjelaskan faktor-faktor ekologis dan spasial yang memengaruhi kejernihan danau.

Baseline & Model

Model LightGBM terbukti lebih baik secara kuantitatif melalui metrik RMSE dan MAE



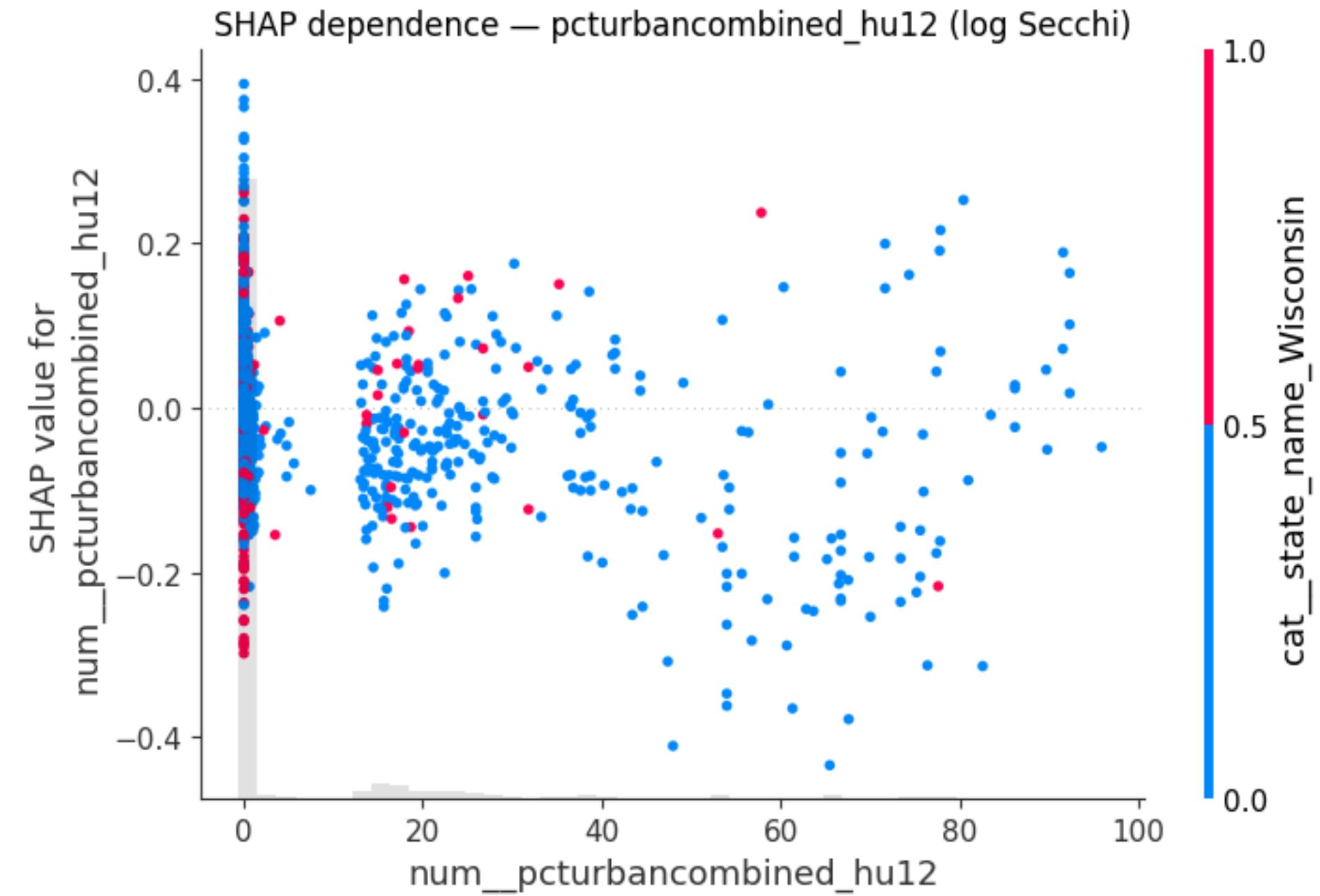
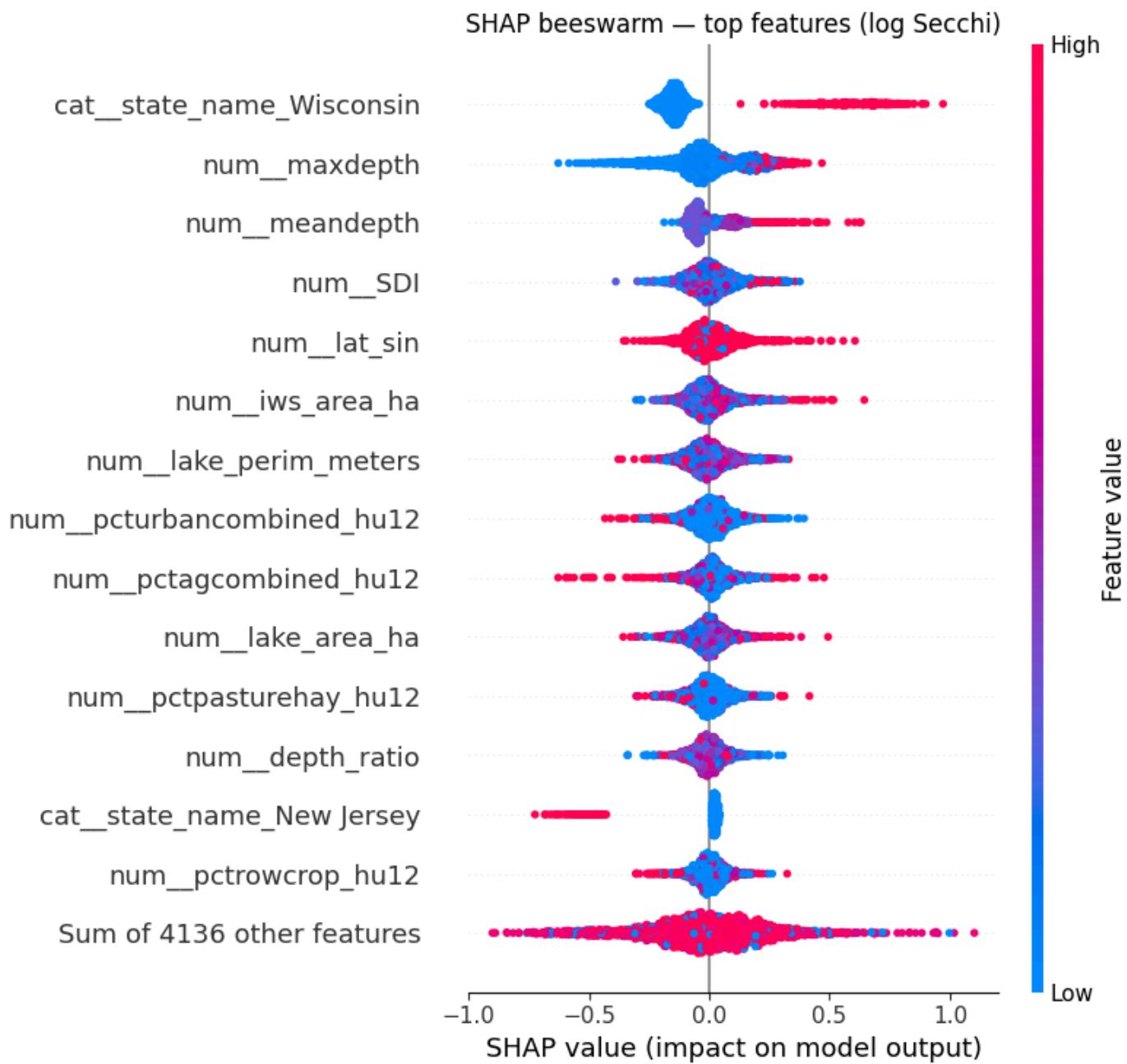
Tabel Metrik

Spatial CV vs Random CV

CV (8-fold)	RMSE_log	MAE_log	Median & Error
Spatial (KMeans)	1.055017	0.823686	115.388373
Spatial (Block Grid)	0.937062	0.792787	66.522391
GroupKFold (HU12)	0.995175	0.748083	55.567471
Random (reference)	1.053665	0.802760	60.882292

“Claim = Spatial CV (lebih ketat)”

SHAP



Penutup

Kesimpulan

- **Ekologi & hayati:** Danau makin keruh saat area urban/pertanian tinggi, danau dangkal paling rentan.
- **Rekreasi & wisata:** Danau dengan Secchi rendah kurang nyaman untuk berenang/berlayar.
- **Pengelolaan & anggaran:** Dengan spatial CV dan interval prediksi, kita bisa memilih prioritas dan pakai dana lebih tepat.

Rekomendasi

- Fokus pemantauan & restorasi di danau dangkal dan wilayah urban tinggi (buffer vegetasi, kurangi run-off).
- **Perbaiki model:** bobot nobs, fitur spasial lebih halus, tambah data nutrien, musim, dan waktu.

Daftar Pustaka

- Soranno PA et al. (2017). LAGOS-NE database.
- Bachmann RW et al. (2017). Secchi depth factors.
- Lundberg SM, Lee SI. SHAP (NIPS 2017).
- Shafer G, Vovk V. Conformal Prediction (JMLR 2008).

Terima Kasih!