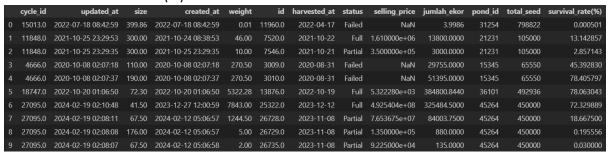
JALA Tech Data Scientist Take Home Test

Faisal Nur Khairudin

- 1. Hasil evaluasi kelengkapan data:
 - a. Tidak terdapat feature description untuk dataset feed tray dan mortalities
 - b. Terdapat *missing values* yang lebih dari 10% pada beberapa kolom di beberapa dataset
 - c. Terdapat beberapa kolom tanggal yang masih memiliki tipe data "object"
 - d. Terdapat negative value pada beberapa kolom yang tidak masuk akal
 - e. Terdapat duplikat data berdasarkan semua kolom terutama pada *feeds dataset* yang memiliki jumlah paling banyak
- 2. Hasil perhitungan SR dan ADG udang di setiap siklus budidaya (10 baris pertama):
 - a. Survival Rate (%)



b. Average Growth Rate/Average Daily Gain (ADG)

	cycle_id	sampled_at	average_weight	ABWt-ABW0	ht-h0	ADG
0	3458.0	2020-04-10	4.37	4.37	34.0	0.128529
1	3458.0	2020-04-17	5.23	0.86	7.0	0.122857
2	3458.0	2020-04-24	6.69	1.46	7.0	0.208571
3	3458.0	2020-05-01	8.21	1.52	7.0	0.217143
4	3458.0	2020-05-08	9.68	1.47	7.0	0.210000
5	3458.0	2020-05-15	11.59	1.91	7.0	0.272857
6	3458.0	2020-05-22	13.82	2.23	7.0	0.318571
7	3458.0	2020-05-29	15.62	1.80	7.0	0.257143
8	3458.0	2020-06-05	18.15	2.53	7.0	0.361429
9	3458.0	2020-06-12	20.49	2.34	7.0	0.334286

- 3. Hasil prediksi model:
 - a. Survival Rate (in percentage) forecast
 - Hasil metrik evaluasi model training:

	Model	MAE	r2_score
0	RandomForestRegressor	4.330758	0.826690
2	XGBRegressor	5.076364	0.803925
4	CatBoostRegressor	6.585183	0.706695
3	LGBMRegressor	7.218221	0.649573
1	AdaBoostRegressor	17.688743	-0.297745

- Hasil *cross validation*:

	MAE	r2	Adj-r2	std-MAE
Model				
CatBoostRegressor	9.379899	0.351475	0.350068	0.327404
RandomForestRegressor	9.513719	0.284244	0.282692	0.333111
LGBMRegressor	9.524140	0.342994	0.341569	0.344540
XGBRegressor	9.720559	0.265480	0.263887	0.235580
AdaBoostRegressor	16.853372	-0.262010	-0.264747	2.408002

- 4. Kesimpulan dan penjelasan terkait fitur atau variabel apa yang penting dalam membuat prediksi
- 5. Rekomendasi mengenai bagaimana budidaya udang perlu dilakukan untuk mencapat hasil yang optimal:
 - Diharapkan para petambak udang mulai mencatat segala data dengan sistem digital, sehingga semua data dapat tercatat. Kemudian, dari data-data yang telah dikumpulkan, sistem bisa melakukan Analisa sesuai dengan kebutuhan para petambak. Harapannya hasil panen udang bisa lebih optimal.
- 6. Link ke repository (github) ke project:
 - Link Github: https://github.com/fnkhairudin/Assignment_JALA
 - MLflow tracking: https://dagshub.com/fnkhairudin/Assignment_JALA.mlflow