

Modelling and Business Insight

STAGE 3 - Final Project



Modeling

Modeling

```
[ ] from sklearn.linear_model import LogisticRegression
    from sklearn.ensemble import RandomForestClassifier
    from xgboost import XGBClassifier

[ ] models = {
    "logistic": LogisticRegression(solver='sag', max_iter=250),
    "RandomForest" :RandomForestClassifier(random_state=42),
    "XGBoost" : XGBClassifier(),
}
```

Kami memutuskan untuk menerapkan tiga jenis model yang berbeda dalam uji coba kami, dan nantinya kami akan melakukan evaluasi ulang terhadap ketiga model tersebut, dengan metrics yang telah ditentukan. Tiga jenis model yang akan kami gunakan adalah **Logistic Regression, Random Forest, dan XGBoost.**

Modeling (Metrics)

```
[ ] models_precision = {}
models_cm = {}
models_cv_precision_test = {}
models_cv_precision_train = {}

[ ] from sklearn.metrics import confusion_matrix, recall_score, precision_score
from sklearn.model_selection import cross_validate, StratifiedKFold

def evaluate_classifier(model_name, model):
    Y_pred = model.predict(X_test)
    models_precision[model_name] = precision_score(y_test, Y_pred)

    kfold = StratifiedKFold(n_splits=5, shuffle=True)
    precision_cv = cross_validate(model, X_train_resampled, y_train_resampled, cv=kfold, scoring='precision', return_train_score=True)
    models_cv_precision_train[model_name] = str(precision_cv['train_score'].mean())
    models_cv_precision_test[model_name] = str(precision_cv['test_score'].mean())

    models_cm[model_name] = confusion_matrix(y_test, Y_pred)
```

Kami telah memutuskan untuk menggunakan metrik utama, yaitu precision sebagai indikator kinerja utama, namun kami menambahkan dua metrik lain yaitu recall score dan confusion matrix.

- Confusion matrix memberikan gambaran rinci tentang klasifikasi yang benar dan kesalahan yang dibuat oleh model.
- Precision score dan Recall Score mengukur kemampuan model dalam mengidentifikasi semua kasus positif dengan benar, kedua metrik ini membantu mengendalikan false positives dan false negatives.



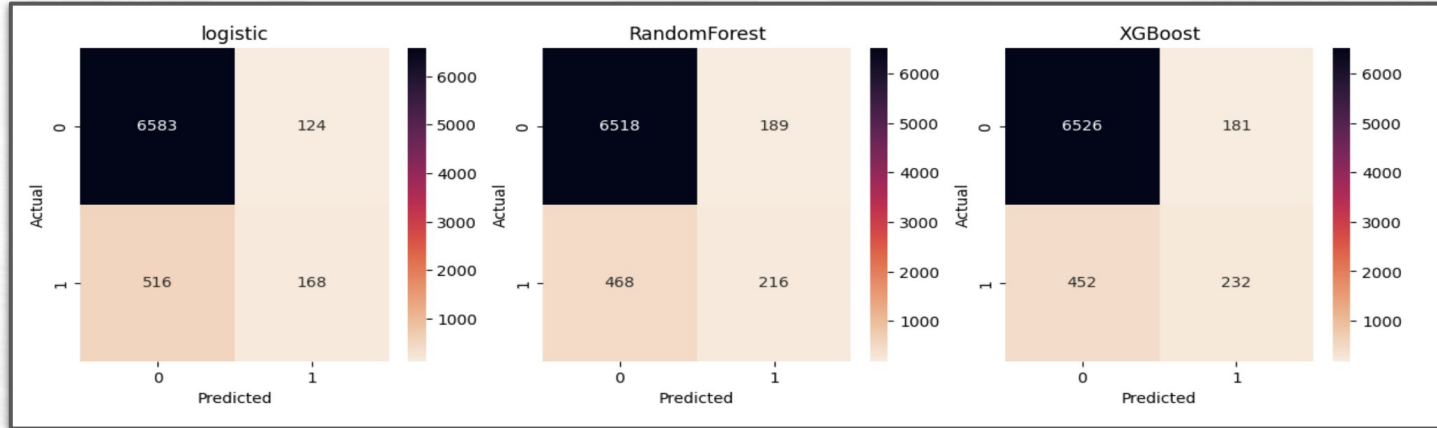
Dengan menggunakan ketiga metrik ini, diharapkan dapat memahami secara menyeluruh kinerja model dan membuat keputusan yang lebih informatif tentang peningkatan atau penyesuaian yang diperlukan.

Modeling (Model Evaluation)

Model	Precision Score	Precision Cross-Val (Train)	Precision Cross-Val (Test)
logistic	0.575342	0.979586	0.979215
RandomForest	0.533333	1	0.968109
XGBoost	0.561743	0.991352	0.968387

Dari ketiga algoritma yang telah dievaluasi, terdapat beberapa observasi penting yang perlu dipertimbangkan. Pertama, dalam hal precision untuk cross-validation (CV), **semua algoritma tidak menunjukkan tanda-tanda overfitting**, yang mengindikasikan bahwa model-model ini konsisten dalam kinerja mereka, khususnya, **Logistic Regression**. Dengan mempertimbangkan nilai precision score yang mencerminkan kemampuan model dalam menggambarkan kasus positif ('y'), Logistic Regression mengungguli dua algoritma lainnya, dengan nilai 0.575.

Modeling (Model Evaluation)



Model logistic regression **memiliki kasus true negative cukup tinggi** daripada kedua model yang lain, walaupun ia **memprediksi true positif yang lebih rendah**. Hal tersebut dikarenakan Logistic Regression seringkali memiliki ambang batas yang lebih konservatif, yang berarti ia lebih cenderung untuk mengklasifikasikan pengamatan sebagai negatif dan berhati-hati saat membuat prediksi positif.

Modeling (Hyperparameter Tuning)

```
lr_clf = LogisticRegression(**best_params)
lr_clf.fit(X_train_resampled, y_train_resampled)
Y_pred = lr_clf.predict(X_test)
precision = precision_score(y_test, Y_pred)

kfold = StratifiedKFold(n_splits=5, shuffle=True)
precision_cv = cross_validate(lr_clf, X_train_resampled, y_train_resampled, cv=kfold, scoring='precision', return_train_score=True)
models_cv_precision_train = str(precision_cv['train_score'].mean())
models_cv_precision_test = str(precision_cv['test_score'].mean())

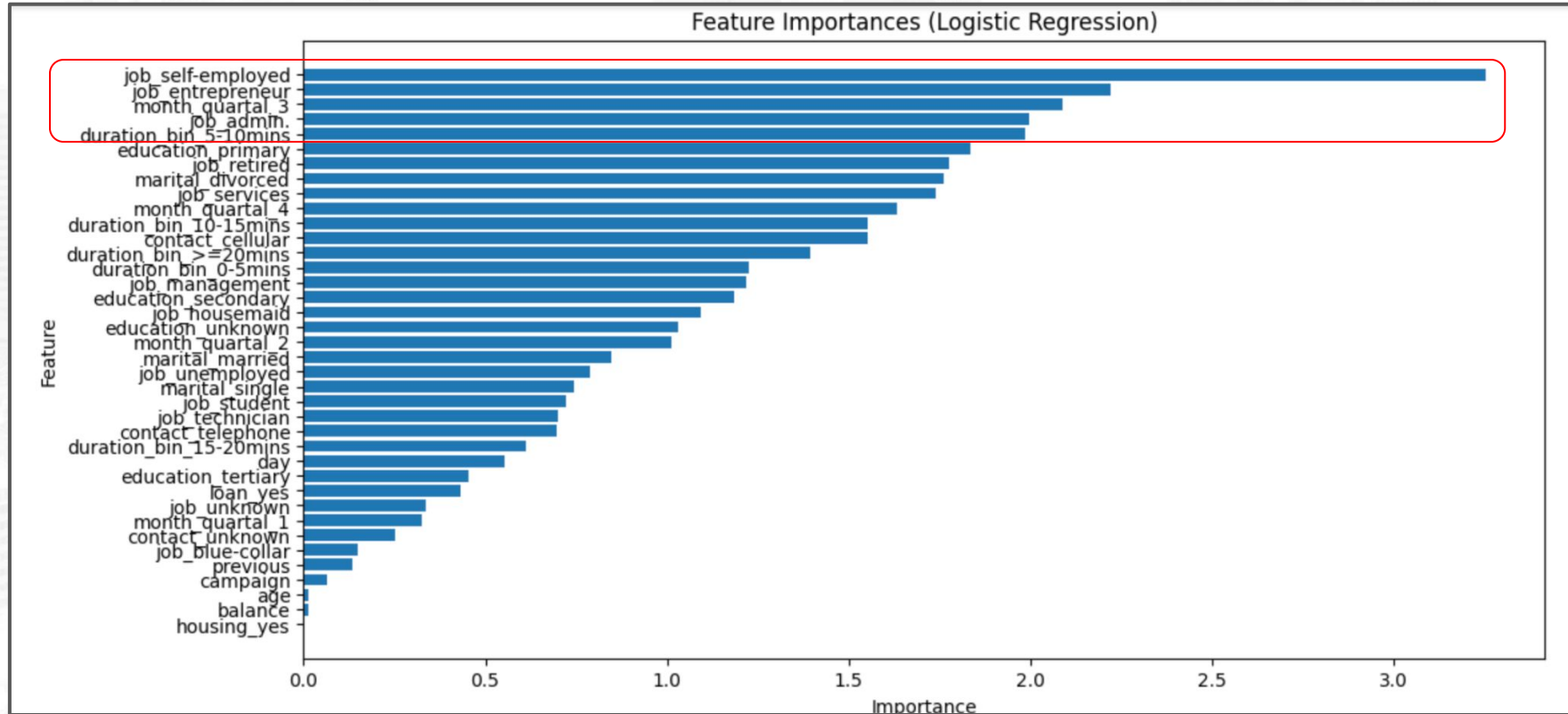
print(f"PRECISION SCORE:\n{precision}")
print(f"CV TRAIN SCORE:\n{models_cv_precision_train}")
print(f"CV TEST SCORE:\n{models_cv_precision_test}")
```

PRECISION SCORE:
0.5738831615120275
CV TRAIN SCORE:
0.9796918922846043
CV TEST SCORE:
0.9795647099771536

Setelah melakukan hyperparameter tuning, nilai precision mengalami sedikit **penurunan**. Oleh karena itu, kami mempertimbangkan untuk kembali menggunakan **model awal** yang memiliki performa lebih baik dalam hal precision.

Feature Importance

Feature Importance



Feature job_self-employed

Dari gambar di slide sebelumnya, fitur job_self-employed (pekerja mandiri) menjadi fitur yang paling penting pada dataset banking ini, dan tingkat kepentingannya unggul jauh dari 37 fitur lainnya di angka lebih dari 3.0 skala importance. Hal ini mungkin disertai beberapa alasan, yaitu:

- **Pekerja mandiri tidak memiliki tunjangan atau jaminan keuangan.** Pekerja mandiri misalnya freelancer cenderung harus membayar pajak dan asuransinya sendiri, sehingga berlangganan term deposit menjadi hal yang harus dilakukan karena ia tidak mendapatkan jaminan apapun.
- **Resiko yang rendah.** Pekerja mandiri biasanya adalah orang yang tipikalnya kurang berani mengambil resiko, karena ia lebih percaya pada dirinya sendiri dalam menjalankan sebuah proyek ketimbang membangun bisnis dengan beberapa rekanan sebagai “entrepreneur”. Deposito berjangka menawarkan resiko yang minim dan cocok untuk pekerja mandiri.
- **Return yang tetap.** Deposito berjangka memiliki return yang cenderung tetap, sangat bagus bagi pekerja mandiri yang penghasilannya tidak tetap dalam rangka kestabilan keuangan.

Feature penting lainnya

Selain `job_self-employed`, fitur-fitur lain seperti `job_entrepreneur`, `month_quartal-3`, `job_admin` dan `duration_bin_5-10mins` juga cukup penting dengan nilai di atas 2 skala importance. Berikut evaluasi tiap fitur.

- **Job_entrepreneur:** pengusaha mungkin memiliki tingkat kemandirian dan pengendalian yang tinggi terhadap keuangan mereka sendiri. Hal ini dapat mempengaruhi kecenderungan mereka untuk berlangganan term deposit sebagai salah satu strategi pengelolaan keuangan yang lebih aman dan stabil serta bisa menjadi dana simpanan untuk jaminan kesehatan dan pensiun.
- **Month_quartal-3:** Mungkin ada faktor musiman atau perubahan kondisi ekonomi yang mempengaruhi keputusan pelanggan pada bulan tersebut. Misalnya, pada kuartal ketiga, mungkin ada pengaruh liburan, peningkatan aktivitas bisnis, atau perubahan siklus keuangan yang dapat mempengaruhi keputusan berlangganan term deposit. Dari sisi bank, banyak penyesuaian dan penentuan keputusan suku bunga pada kuartal 3, pemeriksaan kepatuhan dan regulasi serta persiapan untuk perencanaan keuangan sebelum kuartal 4 yang fokus pada pematangan rencana tahun depan.
- **Job_admin:** Pelanggan dengan pekerjaan administrasi biasanya tipe orang yang terstruktur, terorganisir dan tidak menyukai banyak risiko dan fluktuasi. investasi seperti term deposit sangat cocok, karena berlangganan term deposit menjadi bagian dari strategi pengelolaan keuangan mereka.
- **Duration_bin_5-10mins:** Durasi di bawah 5 menit menjadi durasi mayoritas yang dilakukan yang bisa menunjukkan bahwa kebanyakan nasabah yang tidak berminat tidak akan menerima panggilan lebih dari 5 menit. Durasi 5-10 menit juga memiliki data yang cukup banyak, dimana bisa menjadi titik krusial nasabah yang tertarik kemungkinan besar akan berada pada durasi ini karena ada kesempatan untuk membahas secara lebih rinci manfaat dan keuntungan dari berlangganan term deposit, yang dapat mempengaruhi keputusan pelanggan tanpa bertele-tele.

Business Insight & Recommendation

Business Insight

Berdasarkan feature importance yang kami dapat, berikut merupakan insight bisnis yang kami temukan :

- Nasabah menandatangani uang karena **faktor ketidakstabilan keuangan**, pekerja mandiri (self-employed) seperti freelancer dan entrepreneur **seringkali menghadapi fluktuasi pendapatan** yang lebih tinggi daripada pekerja dengan gaji tetap. **Deposito memberikan jaminan** bahwa sejumlah dana akan tersedia dengan jumlah tertentu pada waktu tertentu, yang dapat membantu mereka mengelola fluktuasi keuangan.
- Tipe pekerja mandiri menjadi tipe orang yang **cenderung loyal dalam investasi deposito berjangka**, karena menjadikan itu sebagai **dana simpanan** layaknya asuransi atau tabungan pensiun bagi karyawan, ini juga bisa berlaku bagi entrepreneur.
- Tipe pekerja mandiri menyukai **resiko yang rendah**, bahkan mereka **rela uangnya ditahan dalam instrumen deposito** (disaat ada banyak instrumen investasi) pada jangka waktu beberapa saat asal uang mereka aman, tipe ini akan **menyimpan deposit dalam waktu yang cukup lama** dengan alasan keamanan.
- Tipe pekerja mandiri mengharapkan **return yang tetap**, dari sini bisa dipelajari **pola likuidasi atau pencarian** deposito mereka karena mereka bisa saja **mencairkan dana secara tidak terduga** tergantung kondisi bisnis pada suatu periode umum.

Business Insight(2)

- **Kuartal 3** menjadi waktu yang tepat bagi konsumen dalam **merencanakan liburan**, berbelanja maupun meningkatkan aktivitas bisnis, ini menjadi **krusial** karena bisa saja banyak dari konsumen **menghindari term deposit**.
- Kuartal 3 juga menjadi waktu untuk **penyesuaian dan evaluasi suku bunga**, bisa jadi ini waktu yang tepat bagi bank untuk **meningkatkan return deposit** untuk meningkatkan minat nasabah.
- **Pekerja admin** menjadi salah satu karyawan yang paling mungkin investasi deposito, hal ini karena ini adalah **tipe orang yang tidak ingin tinggi resiko**, dan biasanya teratur dalam perencanaan keuangan, beberapa dari mereka juga **belum tentu memiliki tunjangan dan jaminan** yang layak, hal ini membuat tipe ini juga potensial.
- Biasanya nasabah yang **tidak tertarik** deposit akan langsung **mengakhiri durasi sebelum 5 menit** pertama, sedangkan menit 5 - 10 sangat krusial sehingga **marketer perlu melatih cara marketing term deposit** dengan kata-kata yang efektif **tidak lebih dari 10 menit**, agar informasi tersampaikan, dan tidak perlu membuang biaya durasi panggilan terlalu banyak.

Recommendation

Berdasarkan insight bisnis yang didapatkan, kami dapat merekomendasikan beberapa hal yang bisa dilakukan dalam menaikkan jumlah nasabah :

- **Memberikan penawaran deposito dengan jangka waktu dan suku bunga khusus**, yang dapat memenuhi kebutuhan bagi orang-orang yang tidak memiliki pendapatan tetap, seperti tabungan yang dikhususkan segmennya bagi mereka yang ingin memiliki tabungan pensiun, asuransi dan jaminan lain yang membantu melindungi mereka dari risiko keuangan yang tidak terduga.
- **Menyediakan layanan konsultasi keuangan** yang sesuai untuk membantu pekerja mandiri dan entrepreneur mengelola keuangan mereka dengan lebih baik. Dengan memahami perlunya stabilitas keuangan, bank dapat membangun hubungan yang kuat dengan kelompok ini dan menarik lebih banyak pelanggan dengan solusi yang relevan.
- **Memberikan pelatihan dan pendidikan keuangan** kepada pekerja mandiri. Hal ini dapat berupa seminar atau kursus online yang membantu mereka memahami bagaimana mengelola uang, membuat anggaran, dan merencanakan masa depan keuangan.
- **Melakukan segmentasi pelanggan berdasarkan durasi panggilan telepon**, dengan fokus pada pelanggan dalam kategori pelanggan dengan durasi 5-10 menit. Dalam pesan pemasaran yang disesuaikan, bank dapat menyoroti manfaat deposito dengan lebih rinci, dan menawarkan penawaran khusus.

Recommendation (2)

- **Melatih marketer dalam marketing deposit 10 menit.** Bank bisa mengurangi biaya durasi telepon dengan melatih marketer mereka menyampaikan informasi yang padat dan menyeluruh tidak lebih dari 10 menit, karena 5-10 menit merupakan waktu krusial, hal ini juga menghindari percakapan yang lebih lama dan memakan waktu yang berujung pada biaya yang membengkak. Pelatihan tim pemasaran menjadi penting untuk berkomunikasi secara efektif dalam waktu singkat.
- **Melakukan review dan penyesuaian suku bunga pada kuartal 3.** Bank perlu melakukan penyesuaian dan kalau bisa peningkatan return pada kuartal 3 melihat bagaimana pola yang terjadi selama 6 bulan sebelumnya. Hal ini juga bisa menarik keinginan nasabah untuk investasi terlebih banyak nasabah menyiapkan dana sisa pada kuartal ini.
- **Menggunakan Data Science untuk mempelajari pola perilaku nasabah.** Bank bisa mempelajari pola investasi nasabah yang telah tersegmentasi dari jenis pekerjaan, pola investasi dan pencairan khususnya pada kuartal 3 dan juga pola komunikasi lewat durasi panggilan di 10 menit pertama, sehingga optimasi bisa dilakukan lagi demi memangkas biaya dan meningkatkan conversion rate.

Thank you