

# Homework Classification



## Estimasi Waktu Pengerjaan

 **3 - 5 jam**

## Jumlah Soal

 **4 Soal**

## Total Point

 **100 poin**

# Teknis Pengerjaan

1. Pekerjaan dilakukan oleh masing-masing individu
2. Menggunakan dataset berikut ini: [Klik](#)
3. File yang perlu dikumpulkan:
  - File **jupyter notebook** (.ipynb) yang berisi source code.
  - File **laporan homework** (.pdf) yang berisi rangkuman dari apa saja yang telah dilakukan.
4. Upload hasil pengerjaan melalui LMS.
  - Masukkan semua file ke dalam **1 file** dengan format **ZIP**.
  - Nama File:  
**Supervised - <Nama>.zip**

# Telecommunication Customers Churn Prediction

Pada task kali ini, student akan diminta untuk melakukan prediksi nih, apakah suatu customer itu akan churn atau tidak. Ini sangat berguna untuk bisa membantu mengidentifikasi customer yang akan churn.

	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
DSL	No	Yes	No	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
DSL	Yes	No	Yes	No	No	No	No	One year	No	Mailed check	56.95	1889.5	No
DSL	Yes	Yes	No	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
DSL	Yes	No	Yes	Yes	No	No	No	One year	No	Bank transfer (autom)	42.3	1840.75	No
Fiber optic	No	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes
Fiber optic	No	No	Yes	No	No	Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	Yes
Fiber optic	No	Yes	No	No	No	Yes	No	Month-to-month	Yes	Credit card (automat	89.1	1949.4	No
DSL	Yes	No	No	No	No	No	No	Month-to-month	No	Mailed check	29.75	301.9	No
Fiber optic	No	No	Yes	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electronic check	104.8	3046.05	Yes
DSL	Yes	Yes	No	No	No	No	No	One year	No	Bank transfer (autom)	56.15	3487.95	No
DSL	Yes	No	No	No	No	No	No	Month-to-month	Yes	Mailed check	49.95	587.45	No
No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automat	18.95	326.8	No



# Definisi masing-masing kolom

- Customerid: id dari customer
- Gender: gender dari customer
- Seniorcitizen: apakah merupakan senior citizen atau tidak
- Partner: apakah memiliki partner atau tidak
- Dependents: apakah memiliki tanggungan atau tidak seperti anak dll
- Tenure: tenure dari langganan customer
- PhoneService: apakah menggunakan layanan phone atau tidak
- MultipleLines: apakah menggunakan multiple lines atau tidak
- InternetService: Tipe dari internet service yang digunakan
- OnlineSecurity: apakah menggunakan fitur online security
- OnlineBackup: apakah menggunakan fitur online backup
- DeviceProtection: apakah menggunakan fitur device protection
- TechSupport: apakah menggunakan fitur tech support atau tidak
- StreamingTV: apakah menggunakan fitur streaming TV atau tidak
- StreamingMovies: apakah menggunakan fitur streaming film atau tidak
- Contract: tipe contract dari customer
- PaperlessBilling: apakah menggunakan fitur paperless billing atau tidak
- PaymentMethod: payment tipe yang digunakan oleh customer
- MonthlyCharges: total charges/biaya bulanan yang dibayarkan
- TotalCharges: total charges secara keseluruhan yang dibayarkan
- Churn: target variabel yang menunjukkan bahwa customer churn atau tidak

# Tugas teman-teman sebagai Data Scientist

- Kerjakan secara berkelompok (team final project)
- (1) Lakukan EDA dan preprocessing sederhana (30 point)
  - Jelaskan Fitur mana yang sebaiknya digunakan dari hasil EDA?
- (2) Lakukan feature engineering (10 point)
  - Apakah ada feature tambahan lain yang mendukung? Jelaskan mengapa menggunakan feature tersebut.

Referensi tambahan:

- Feature selection: [Google](#)
- Feature engineering terkait date: <https://towardsdatascience.com/feature-engineering-on-date-time-data-90f6e954e6b8>

# Tugas teman-teman sebagai Data Scientist

- (3) Lakukan training model & **prediksi churn** sebagai variabel target (30 point)
  - Pastikan menggunakan lebih dari 1 model dan harus ada model dengan tipe ensemble model seperti random forest, XGBoost dll (20 point)
  - Lakukan tuning hyperparameter, cari mana model yang paling baik (10 point)
- (4) Evaluasi model dengan metrics Recall dan ROC-AUC (30 point)
  - Jelaskan & berikan analisis mengapa memilih model tersebut sebagai model akhir yang digunakan. (10)
  - Jelaskan rekomendasi bisnis yang tepat (20)

Referensi tambahan:

- Feature selection: [Google](#)
- Feature engineering terkait date:

<https://towardsdatascience.com/feature-engineering-on-date-time-data-90f6e954e6b8>

# Submission

- Submit berupa file notebook .ipynb dan document report .pdf
- Template report dapat dilihat [disini](#)