

## Есептер

### I. Теориялық тапсырмалар

#### 1-тапсырма. Метрикаларды түсіндіру

Берілген confusion matrix үшін келесі метрикаларды есептеңіз:

	Нақты +	Нақты –
Болжам +	45	15
Болжам –	5	435

#### Сұрақтар:

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. Қай метрика бұл жағдайда ең маңызды және неге?

#### 2-тапсырма. Accuracy неге жеткіліксіз?

Берілген жағдайды талдаңыз:

- Dataset: 10 000 үлгі
- Positive class: 200 (2%)
- Model барлық объектілерді **negative** деп болжайды

#### Сұрақтар:

1. Accuracy неге тең?
2. Бұл модельді «жақсы» деп атауға бола ма?
3. Қандай метрикалар қолдану керек?

### 3-тапсырма. Кросс-валидация теориясы

1. K-fold cross-validation деген не?
2. Leave-One-Out CV қашан тиімсіз?
3. Stratified K-Fold не үшін қажет?
4. Overfitting-ті CV қалай анықтайды?

### 4-тапсырма. Гиперпараметрлер

Келесі параметрлерді **параметр** немесе **гиперпараметр** деп жіктеңіз:

Элемент	Түрі
Learning rate	
Model weights	
Number of trees	
Regularization coefficient	
Bias	

## 5-тапсырма. SMOTE теориясы

1. SMOTE алгоритмінің негізгі идеясы
2. Формуласы қандай?
3. Қандай жағдайда SMOTE қауіпті?
4. SMOTE-ті train мен test-ке бірдей қолдануға бола ма?

## 6-тапсырма. Class weight

1. `class_weight="balanced"` нені өзгертеді?
2. Loss функциясына қалай әсер етеді?
3. SMOTE пен class\_weight айырмашылығы неде?

## 7-тапсырма. Оптимизация

1. Gradient Descent формуласы
2. Learning rate үлкен болса не болады?
3. L1 және L2 регуляризацияның айырмашылығы
4. Қайсысы feature selection жасайды?

## II. Практикалық тапсырмалар (Python / Colab)

Ескерту: барлық тапсырмалар алдыңғы Colab кодына негізделген

## 8-тапсырма. Базалық модель

1. SMOTE пен class\_weight-ті алып тастаңыз
2. Logistic Regression үйретіңіз
3. Accuracy, Recall, F1 салыстырыңыз
4. Қорытынды жазыңыз



## 9-тапсырма. SMOTE әсері

1. Модельді SMOTE-сыз үйретіңіз
2. Модельді SMOTE-пен үйретіңіз
3. Recall және ROC-AUC салыстырыңыз
4. Қайсысы minority class үшін жақсы?

## 10-тапсырма. Class weight әсері

1. `class_weight=None`
2. `class_weight="balanced"`
3. F1-score өзгерісін салыстырыңыз
4. Қай жағдайда class\_weight жеткілікті?

## 11-тапсырма. Гиперпараметрлерді іріктеу

1. GridSearchCV арқылы `C` және `penalty` таңдаңыз
2. `scoring="f1"` және `scoring="roc_auc"` салыстырыңыз
3. Нәтижелер айырмашылығын түсіндіріңіз

## 12-тапсырма. Кросс-валидация қателігі

1. 3-fold және 10-fold CV қолданыңыз
2. Орташа F1-score салыстырыңыз
3. Variance қалай өзгереді?

### 13-тапсырма. Метрика таңдау

Берілген сценарийлер үшін ең дұрыс метриканы таңдаңыз:

1. Алаяқтықты анықтау
2. Спам-фильтр
3. Медициналық диагноз
4. Студенттің оқудан шығуын болжау

Әрқайсысына неге деген түсініктеме беріңіз.

## III. Зерттеу / ойлану тапсырмалары (Advanced)

### 14-тапсырма. Leakage

1. Data leakage деген не?
2. SMOTE leakage қалай туады?
3. Pipeline оны қалай шешеді?

## 15-тапсырма. Threshold optimization

1. Неге 0.5 әрқашан дұрыс емес?
2. Threshold-ты қалай оптимизациялауға болады?
3. Recall-ды максимизациялау үшін threshold қалай өзгереді?

## 16-тапсырма. Реал кейс (эссе)

### Тақырып:

«Теңгерімсіз деректермен жұмыс істегенде метрика мен оптимизацияның рөлі»

Көлемі: 1–2 бет

### Құрылымы:

- Мәселе
- Қолданылған әдістер
- Қорытынды