

Есептер

● 1-деңгей. Түсінуді тексеру (теория)

Есеп 1. Анықтамалар

Қысқаша жауап беріңіз:

1. Ансамбль әдісі деген не?
2. Неліктен бірнеше модель бір модельден жақсы жұмыс істеуі мүмкін?
3. Bias пен Variance деген не?

Есеп 2. Bagging идеясы

Түсіндіріңіз:

1. Bootstrap sample деген не?
2. Bagging қандай проблеманы азайтады?
3. Bagging параллель орындала ма? Неге?

Есеп 3. Random Forest ерекшелігі

Төмендегілердің қайсысы **Random Forest-ке** тән, қайсысы емес — белгілеңіз:

- Bootstrap sampling
- Feature randomness
- Sequential learning
- Decision trees
- Gradient optimization

● 2-деңгей. Есептеу және логика

Есеп 4. Majority voting

3 классификатор келесі болжамдарды берді:

Модель	Болжам
M1	1

M2	0
M3	1

Тапсырма:

1. Ансамбль болжамы қандай?
2. Егер бір модельдің салмағы 2 есе үлкен болса, нәтиже өзгере ме?

● **3-деңгей. Bagging & Random Forest (практика)**

Есеп 6. Bagging (Python)

Берілген:

```
from sklearn.ensemble import BaggingClassifier  
from sklearn.tree import DecisionTreeClassifier
```

Тапсырма:

1. Base model ретінде Decision Tree қолданыңыз
2. `n_estimators = 30`
3. Accuracy есептеңіз
4. Нәтижені бір ағашпен салыстырыңыз

Есеп 7. Random Forest гиперпараметрлері

Random Forest үшін:

1. `n_estimators` артса не болады?
2. `max_depth` азайса модельге әсері?
3. Неліктен Random Forest overfitting-ке төзімді?

● **4-деңгей. Gradient Boosting**

Есеп 8. Бустинг логикасы

Түсіндіріңіз:

1. Неліктен Gradient Boosting sequential?
2. Қалдықтар (residuals) деген не?
3. Learning rate-тің рөлі қандай?

Есеп 9. Gradient Boosting (Python)

Тапсырма:

1. GradientBoostingClassifier қолданыңыз
2. `n_estimators = 50, 100, 200`
3. Accuracy өзгерісін салыстырыңыз
4. Қорытынды жасаңыз

● 5-деңгей. Stacking

Есеп 10. Stacking құрылымы

Берілген модельдер:

- Decision Tree
- Random Forest
- SVM

Тапсырма:

1. Қайсысы Level-0 модель?
2. Мета-модель қандай болуы мүмкін?
3. Неліктен train деректерді тікелей қолдануға болмайды?

Есеп 11. Stacking (Python)

1. Кемінде 2 базалық модель қолданыңыз
2. Logistic Regression — meta-model
3. Accuracy есептеңіз
4. Random Forest-пен салыстырыңыз

● 6-деңгей. Аналитикалық

Есеп 12. Bias–Variance талдауы

Келесі әдістер үшін Bias пен Variance-ті салыстырыңыз:

1. Single Decision Tree
2. Bagging
3. Random Forest
4. Gradient Boosting

Есеп 13. Қай әдісті таңдау керек?

Қай әдісті таңдар едіңіз және неге:

1. Шағын датасет
2. Үлкен noisy деректер
3. Feature саны өте көп
4. Ең жоғары ассурасу қажет

★ Қорытынды есеп (Mini-project)

Есеп 14. Толық салыстыру

Бір датасетте:

- Bagging
- Random Forest
- Gradient Boosting
- Stacking

Тапсырма:

1. Барлығын үйретіңіз
2. Ассурасу салыстырыңыз
3. Қайсысы ең тұрақты?
4. Қайсысы ең баяу?

🎓 Теория

1. Неліктен Random Forest feature importance бере алады?
2. Gradient Boosting неге overfitting жасай алады?
3. Stacking қашан тиімді емес?
4. Bagging bias-ты азайта ма?