

به نام خدا

مطالب توجیهی پروژه‌ی چهارم

هوش مصنوعی - بهار ۱۳۹۹

با سلام. همانطور که در جلسه‌ی توجیهی ۲۷ اردیبهشت بیان شد، مطالب زیر برای شفاف سازی در اختیار شما قرار می‌گیرد :

1. در دادگان پروژه، جزئیاتی همچون مقادیر منفی در ستون‌های مقدار و هزینه و همچنین مقدار صفر در ستون تعداد کالا وجود دارد. برای سادگی می‌توانید از این موضوع صرف‌نظر کنید (می‌توانید فرض کنید این دادگان واقعی هستند و واکنشی به آن‌ها نشان ندهید و مدل‌ها را با وجود آن‌ها آموزش دهید). هر رفتاری با این جزئیات مقبول است.
2. در مرحله‌ی پیش‌پردازش، انجام موارد زیر اجباریست :
 - a. تحلیل تاریخ (به هر روش معناداری)
 - b. برخورد با دادگان categorical (ستون کشور) متناسب با مدل. د رمورد تفاوت Label Encoding و One Hot Encoding توضیح دهید (در مورد اینکه چرا در برخی مواقع نمی‌توان از یکی استفاده کرد) و معایب دو روش را مطرح کنید.
 - c. اجرای scale بر روی دادگان برای مدل‌هایی که نیاز دارند. توضیح دهید چرا عمل scale به این مدل‌ها کمک می‌کند.
3. برای مرحله‌ی **پیش‌پردازش**¹ می‌توانید از لینک‌های زیر استفاده کنید :
 - a. **برای تحلیل تاریخ**²
 - b. **برای Scale دادگان**³
 - c. **برای One Hot Encoding**⁴

¹ <https://scikit-learn.org/stable/modules/preprocessing.html>

²

<https://medium.com/@swethalakshmanan14/simple-ways-to-extract-features-from-date-variable-using-pythhon-60c33e3b0501>

³ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

⁴ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>

d. برای [Label Encoding](#)⁵

e. برای [تقسیم دادگان](#)⁶

f. برای [محاسبه‌ی Information Gain](#)⁷

4. برای [رده‌بندها](#)⁸ی ابتدایی می‌توانید از لینک‌های زیر استفاده کنید :

a. برای [ارزیابی و اعتبارسنجی](#)⁹

b. برای [درخت تصمیم](#)¹⁰

c. برای [KNN](#)¹¹

d. برای [Logistic Regression](#)¹²

5. برای روش‌های [ensemble](#)¹³ می‌توانید از لینک‌های زیر استفاده کنید :

a. برای [Bagging](#)¹⁴

b. برای [جنگل تصادفی](#)¹⁵

c. برای [Voting](#)¹⁶

برای توضیحات بیشتر می‌توانید ویدیوی جلسه‌ی توجیهی را که در صفحه درس و همچنین در درگاه [آپارات](#)¹⁷ بارگذاری شده را مشاهده کنید.

5

<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html#sklearn.preprocessing.LabelEncoder>

⁶ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

⁷ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html

⁸ <https://stackoverflow.com/questions/41844311/list-of-all-classification-algorithms>

⁹ <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

¹⁰ <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

¹¹ <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

¹² https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

¹³ <https://scikit-learn.org/stable/modules/ensemble.html>

¹⁴ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html>

¹⁵ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

¹⁶ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html>

¹⁷ <https://www.aparat.com/v/2EnAv>