نماتک
آموزش صحیح

بسته:

داده کاوی با پایتون

فصل اول

- معرفی علم داده
- معرفی مراحل علم داده
- معرفی مرحله جمع‌آوری داده
- معرفی مرحله پیش‌پردازش داده
- معرفی تحلیل اکتشافی داده
- معرفی مراحل اجرای الگوریتم ماشین لرنینگ
- معرفی دو دسته کلی الگوریتم‌های ماشین لرنینگ
- معرفی کاربردهای علم داده
- معرفی داده‌های ساختاریافته
- معرفی داده‌های بدون ساختار

نماتک
آموزش صحیح

فصل دوم

- معرفی روش‌های مدیریت missing value
- معرفی انواع outliers
- معرفی روش‌های شناسایی outliers
- معرفی روش‌های مدیریت outliers
- معرفی روش استانداردسازی
- معرفی روش نرمال‌سازی
- معرفی خلاصه‌سازی دیتا
- معرفی مهندسی ویژگی
- معرفی data encoding
- تشریح نحوه Categorical Encoding
- معرفی مبحث اعتبارسنجی دیتا
- معرفی حکمرانی داده
- معرفی متادیتا
- معرفی ابزارها و تکنولوژی‌های آماده‌سازی داده‌ها
- معرفی مفهوم کتابخانه و پکیج
- معرفی کتابخانه numpy
- معرفی نحوه اضافه کردن کتابخانه numpy به اسکریپت
- معرفی تابع abs و sqrt از کتابخانه numpy
- معرفی تابع random.normal و random.uniform از کتابخانه numpy
- معرفی تابع min, max, mean, median, std از کتابخانه numpy

- معرفی تابع seed از کتابخانه numpy
- معرفی نحوه تبدیل نوع داده
- تشریح تفاوت عملیات جمع در list و numpy.ndarray
- معرفی shape از کتابخانه numpy
- معرفی تابع sum و round و subtract از کتابخانه numpy
- معرفی تابع divide و floor_divide از کتابخانه numpy
- معرفی تابع arange و reshape از کتابخانه numpy
- معرفی تابع unique از کتابخانه numpy
- معرفی تابع append و concatenate از کتابخانه numpy
- تشریح مزیت استفاده از میانه نسبت به سایر پارامترها، در مجموعه‌داده‌هایی که مشکوک به وجود نویز در آن هستیم
- مصورسازی (visualization)
- معرفی کتابخانه matplotlib
- معرفی نحوه اضافه کردن کتابخانه numpy به اسکریپت
- معرفی نمودار hist از کتابخانه matplotlib
- معرفی نمودار plot از کتابخانه matplotlib
- معرفی تابع xlabel و ylabel و legend از کتابخانه matplotlib
- معرفی تابع fill_between از کتابخانه matplotlib
- معرفی تابع grid از کتابخانه matplotlib
- معرفی تابع subplot از کتابخانه matplotlib
- معرفی نمودار bar و barh از کتابخانه matplotlib
- معرفی تابع bar, barh, xticks, title از کتابخانه matplotlib

- معرفی تابع savefig از کتابخانه matplotlib
- معرفی نمودار pie از کتابخانه matplotlib
- معرفی نمودار scatter از کتابخانه matplotlib
- معرفی تابع figure از کتابخانه matplotlib
- معرفی rcparams از کتابخانه matplotlib
- تشریح نحوه ترسیم نمودار scatter به‌صورت سه‌بعدی
- معرفی نمودار boxplot از کتابخانه matplotlib
- معرفی کتابخانه‌های پایتون جهت مصورسازی
- معرفی کتابخانه pandas
- معرفی تابع read_csv از کتابخانه pandas جهت خواندن فایل csv
- معرفی متد head و tail
- معرفی توابع آماری بر روی دیتافریم‌ها
- معرفی متد describe
- معرفی نحوه اشاره به یک ستون خاص در پانداس
- معرفی اتربیوت dtypes
- معرفی تابع read_excel از کتابخانه pandas جهت خواندن فایل excel
- معرفی متد Value_counts
- معرفی دستور iloc
- معرفی مفهوم ماسک کردن
- معرفی نحوه خواندن نام ستون‌ها و سطرها در پانداس
- معرفی دستور loc و تفاوت آن با دستور iloc
- معرفی تابع cut از کتابخانه pandas

- معرفی نحوه کپی گرفتن از یک دیتافریم
- معرفی تابع collect از پکیج gc جهت پاک‌سازی کامل حافظه
- معرفی متد info
- معرفی مجیک کامند
- معرفی متد select_dtypes
- معرفی متد sort_values
- معرفی نحوه نوشتن یک دیتافریم بر روی فایل اکسل
- معرفی نحوه تغییر نام ستون‌های یک دیتافریم
- نحوه پیدا کردن و هندل کردن missing value در هر ستون از دیتافریم
- نحوه نمونه‌گیری از دیتافریم
- تشریح مبحث data integration
- معرفی نحوه ساخت DataFrame
- معرفی متد reset_index
- معرفی متد merge
- معرفی متد concat
- معرفی متد groupby و agg
- معرفی متد rename
- معرفی متد set_index
- تشریح مثالی در خصوص درصد ریزش مشتری در هر منطقه
- معرفی راحل اعمال دو فانکشن aggregate بر روی یک ستون
- معرفی متد drop
- معرفی متد drop_duplicates

- تشریح مثالی در خصوص یافتن درصد ریزش مشتری در هر منطقه
- تشریح مثالی در خصوص یافتن سطرهای حذف شده در دستور drop_duplicates
- تشریح نحوه شناسایی داده‌های پرت
- معرفی فرآیند استفاده از پکیج sklearn
- تشریح نحوه استفاده از MinMaxScaler برای پیش‌پردازش داده‌ها
- تشریح نحوه استفاده از StandardScaler برای پیش‌پردازش داده‌ها
- معرفی متد inverse_transform جهت دی نرمال کردن دیتا
- معرفی الگوریتم PCA و استفاده آن جهت کاهش ابعاد دیتا
- معرفی ویژگی explained_variance_ratio_
- معرفی تابع LabelEncoder و جهت تبدیل ستون‌های غیرعددی به عدد
- تشریح مثالی در خصوص تبدیل تمامی ستون‌های غیرعددی به عدد
- تشریح تبدیل تمامی ستون‌های غیرعددی به عدد با استفاده از get_dummies در پکیج پانداس
- معرفی الگوریتم KNNImputer جهت هندل کردن داده‌های miss شده
- معرفی بررسی پرت بودن داده‌ها با استفاده از الگوریتم LocalOutlierFactor
- معرفی چالش‌های ناشی از داده‌های نامتوازن
- معرفی overfitting
- معرفی Low recall
- معرفی انواع داده‌های نامتوازن
- تشریح مدیریت داده‌های نامتوازن
- معرفی تکنیک‌های Resampling
- معرفی الگوریتم Near Miss

- معرفی الگوریتم Tomek Link
- معرفی الگوریتم SMOTE
- معرفی Cost-Sensitive Learning
- معرفی روش‌های ترکیبی (Ensemble Methods)
- معرفی متریک‌های ارزیابی داده‌های نامتوازن
- معرفی الگوریتم بالانسینگ RandomUnderSampler از پکیج imblearn
- معرفی الگوریتم بالانسینگ RandomOverSampler از پکیج imblearn
- معرفی الگوریتم بالانسینگ TomekLinks از پکیج imblearn
- معرفی الگوریتم بالانسینگ SMOTE از پکیج imblearn
- معرفی الگوریتم بالانسینگ NearMiss از پکیج imblearn

فصل سوم

- تعریف یادگیری نظارت شده
- معرفی مسیر یادگیری نظارت شده
- معرفی دو دسته کلی یادگیری نظارت شده
- معرفی الگوریتم‌های کلیدی در یادگیری نظارت شده
- معرفی الگوریتم k-Nearest Neighbors
- معرفی الگوریتم Linear Regression
- معرفی الگوریتم Logistic Regression
- معرفی الگوریتم Decision Tree
- معرفی مفهوم underfitting و overfitting
- معرفی مزایای درخت تصمیم
- معرفی الگوریتم Support Vector Machines
- معرفی مفاهیم اساسی در الگوریتم Support Vector Machines
- معرفی مفهوم support vector و margin
- معرفی الگوریتم Artificial Neural Networks
- معرفی پارامترهای الگوریتم شبکه عصبی
- معرفی رویکردهایی جهت جلوگیری از overfitting
- معرفی انواع متریک‌های ارزیابی Classification
- معرفی accuracy
- معرفی recall
- معرفی precision

- معرفی F1
- معرفی متریک‌های ارزیابی Regression
- معرفی Mean Squared Error
- معرفی Root Mean Squared Error
- معرفی Mean Absolute Error
- معرفی Mean Squared Percentage Error
- معرفی استراتژی‌های ارزیابی
- معرفی استراتژی Train-Test Split
- معرفی استراتژی K-Fold Cross-Validation
- معرفی استراتژی Stratified Sampling
- مدل‌سازی و ارزیابی با رویکرد Classification
- معرفی استراتژی ارزیابی train_test_split
- معرفی الگوریتم GaussianNB
- معرفی الگوریتم‌های صحت سنجی
- معرفی الگوریتم MultinomialNB
- معرفی الگوریتم KNeighborsClassifier
- معرفی الگوریتم DecisionTreeClassifier
- معرفی متد predict_proba الگوریتم‌های پیش‌بینی
- معرفی تابع export_graphviz
- تشریح نحوه ترسیم خروجی تابع export_graphviz با استفاده از پکیج graphviz
- معرفی تابع export_text
- معرفی الگوریتم DecisionTreeClassifier

- معرفی الگوریتم MLPClassifier
- معرفی الگوریتم SVC
- معرفی الگوریتم LogisticRegression
- معرفی استراتژی ارزیابی cross_val_predict و cross_val_score
- معرفی استراتژی ارزیابی KFold
- معرفی استراتژی ارزیابی StratifiedKFold
- حل نمونه مسئله کلاسیفیکیشن گل‌های وحشی در سایت دیتا کوئیز
- معرفی روش‌های بهینه‌سازی پارامترها
- معرفی روش Grid Search
- معرفی روش Random Search
- معرفی روش Bayesian Search
- معرفی تابع make_scorer جهت ایجاد متریک‌های ارزیابی شخصی
- حل مسئله پیش‌بینی نمره ریاضی دانش‌آموزان در سایت دیتا کوئیز

فصل چهارم

- تعریف یادگیری بدون نظارت
- معرفی دو دسته کلی یادگیری بدون نظارت
- معرفی الگوریتم‌های خوشه‌بندی
- معرفی الگوریتم K-Means
- معرفی الگوریتم DBSCAN
- معرفی الگوریتم‌های پترن‌های پر تکرار
- معرفی الگوریتم‌های پترن‌های پر تکرار
- معرفی قواعد انجمنی
- معرفی کاربردهای یادگیری نظارت نشده
- معرفی انواع متریک‌های ارزیابی Clustering
- معرفی Silhouette Score
- معرفی Davies-Bouldin Index
- معرفی Adjusted Rand Index (ARI)
- خوشه‌بندی با استفاده از کتابخانه sklearn
- خواندن دیتای بیماران دیابتی و نرمال‌سازی دیتا با استفاده از الگوریتم MinMaxScaler
- استفاده از الگوریتم Kmeans برای خوش بندی دیتا
- تحلیل خریدهای مشتری‌ها بر اساس روش RFM
- مشاهده مراکز دیتاهای خوشه‌بندی شده
- مشاهده محل قرارگیری هر دیتا بر اساس خوشه

- ارزیابی مدل خوشه‌بندی با استفاده از معیار ارزیابی silhouette_score
- معرفی دو روش جهت پیدا کردن تعداد بهینه خوشه
- استفاده از الگوریتم DBSCAN برای خوش بندی دیتا
- الگوهای پر تکرار و قواعد انجمنی با استفاده از پکیج mlxtend
- معرفی الگوریتم apriori
- معرفی الگوریتم fpgrowth
- معرفی تابع association_rules
- معرفی کلاس TransactionEncoder جهت پیش‌پردازش داده‌ها

فصل پنجم

- روش‌های ترکیبی (Ensemble Methods)
- معرفی روش‌های پایه
- معرفی روش Voting
- معرفی روش Averaging
- معرفی روش‌های پیشرفته
- معرفی روش Bagging
- معرفی روش Random Forest
- معرفی روش Stacking
- معرفی روش Boosting
- معرفی الگوریتم‌های بر پایه Boosting
- معرفی پیاده‌سازی الگوریتم EnsembleVoteClassifier از پکیج mlxtend
- معرفی پیاده‌سازی الگوریتم BaggingClassifier از پکیج sklearn
- معرفی پیاده‌سازی الگوریتم RandomForestClassifier از پکیج sklearn
- معرفی پیاده‌سازی الگوریتم AdaBoostClassifier از پکیج sklearn
- معرفی پیاده‌سازی الگوریتم StackingCVClassifier از پکیج sklearn
- سه الگوریتم قدرتمند بر پایه gradient boosting machine
- معرفی پیاده‌سازی الگوریتم XGBClassifier از پکیج xgboost
- معرفی پیاده‌سازی الگوریتم LGBMClassifier از پکیج lightgbm
- معرفی پیاده‌سازی الگوریتم CatBoostClassifier از پکیج catboost

فصل ششم

- معرفی سایت Kaggle
- معرفی اهداف مهندسی ویژگی
- اهمیت ساخت ویژگی‌های جدید از ویژگی‌های فعلی دیتا
- معرفی Mutual Information
- معرفی روش‌های ساخت ویژگی‌های جدید
- معرفی Building-Up and Breaking-Down Feathres
- معرفی روش تحلیل اساسی (PCA)
- معرفی روش Target Encoding
- پیاده‌سازی الگوریتم TargetEncoding با استفاده از پکیج category_encoders
- پیاده‌سازی الگوریتم CatBoostEncoder با استفاده از پکیج category_encoders
- پیاده‌سازی الگوریتم RFE(recursive feature elimination) با استفاده از پکیج sklearn
- معرفی روش Permutation Importance
- تشریح مثالی از یک مسئله رگرسیون از سایت Kaggle