

データサイエンス 基礎

第 2 回 効果測定 問題

問題 1 配点: 1 問 4 点、40 点

キノコのデータ群である「mushroom-test.csv」を読み込み、食べられるキノコか毒キノコかを判別するモデルを作成します。それぞれのコメント欄をよく読み、適切なソースコードを記載して、ipynb ファイルを提出してください。

提出するファイル名は、「ds_exam2_1_氏名.ipynb」とすること。

- ① 読み込んだデータの先頭 10 行を表示する
- ② mushroom データの行数、列数を表示する
- ③ mushroom データ内の欠損値の数を表示する
- ④ 読み込んだデータの説明変数をすべて One-Hot Encoding し、mushroom_dummy という変数名の新しい DataFrame に格納する
- ⑤ y に目的変数の列を代入する
- ⑥ X に説明変数の列を代入する
- ⑦ 学習データとテストデータに分割する (test_size = 0.2 , random_state = 0 とすること)
- ⑧ モデルの初期化、学習を行う
- ⑨ 学習データの正解率の出力
- ⑩ テストデータで予測を行う

データサイエンス 基礎

第 2 回 効果測定 問題

問題 2 配点:40 点

画像データ(ds_koukasokutei2-2.jpg)を読み込み、以下の処理を行う「ipynb」ファイルと、画像操作を行った「ds_koukasokutei2-2 氏名.jpg」を提出してください。

提出するソースコードを記載したファイル名は「ds_exam2_2_氏名.ipynb」とすること。

- ① 「ds_koukasokutei2-3.jpg」を読み込み、実行環境上で画像を表示する。
- ② shape を使ってデータの形状を出力する
- ③ 画像データのサイズを取得し、元データの 4 分の 1 の大きさに縮小し、縮小後の画像を表示する。
- ④ ③で縮小した画像データを左に 90 度回転させ、回転後の画像を表示する。
- ⑤ ④の画像データのうち、料理部分だけを四角形に切り取り、切り取った画像を表示する
- ⑥ ⑤の画像データを 15×15 で平滑化する。
- ⑦ ⑥の画像データのうち、自分自身の最も好きな具材が表示されている部分を四角形で囲う。四角形の色は、見やすい色にしてください。
- ⑧ ⑦までの処理内容を保存するために、編集した画像データを「ds_koukasokutei2-3 氏名.jpg」という名前で保存する。

実装においては、下記を考慮してください。

- コメントを適切に記載する。
- 変数名、関数名などはプログラムに適した名称とする。

問題 3 配点:20 点

とあるホテルの宿泊者に対するアンケート結果をまとめた「アンケート.txt」を元に、アンケートの概要がわかるワードクラウドを作成し、作成したワードクラウドと、画像データを作成するソースコードを提出してください。

- ストップワードを使用する場合は、配付した「stop_words.txt」を用いること。ストップワードに追記した場合はそのテキストファイルも提出すること。
- 文章中の不要な単語「●」や「—————」を削除するなど、文章の特徴と関係ない単語を削除すること。
- 提出するファイルのファイル名は「ds_exam2_3_氏名.png」、「ds_exam2_3_氏名.ipynb」とすること。

実装においては、下記を考慮してください。

- コメントを適切に記載する。
- 変数名、関数名などはプログラムに適した名称とする。