

## سوال هفتم

نمونه داده‌های فایل US Presidential Data را باز کنید. این فایل حاوی آمار برد و باخت انتخابات ریاست جمهوری آمریکا است که دارای دو دسته و ۱۳ ویژگی است. دو دسته، شامل برد (۱) و باخت (۰) در انتخابات ریاست جمهوری می‌شود. ۱۳ ویژگی شامل موارد زیر در متن‌های سخنرانی انتخاباتی است:

(۱) نسبت کلمات نشان دهنده:

- خوش بینی
- بدبینی
- گذشته
- حال
- آینده

(۲) تعداد دفعاتی که نامزد از حزب خود نام می‌برد

(۳) تعداد دفعاتی که نامزدی از حزب رقیب نام می‌برد

(۴) معیاری از محتوای نشان دهنده:

- صراحت
- وجدان
- برون‌گرایی
- موافق بودن
- عصبیت
- احساسی بودن

داده‌ها را به مجموعه Train و Test تقسیم کنید.

(a) داده‌های مجموعه Train را با الگوریتم NN-1 دسته‌بندی کنید. خطای دسته‌بندی بر روی داده Train را گزارش کنید.

(b) بخش ۱ را با الگوریتم‌های K-NN با مقادیر مختلف برای K امتحان کنید. نمودار خطای الگوریتم را بر اساس k‌های مختلف به دست آورید. تاثیر مقدار k را در دقت الگوریتم K-NN چگونه ارزیابی می‌کنید.