

گزارش کار : دسته‌بندی داده‌های بانک با الگوریتم Naive Bayes

ارائه به : دکتر محسن عباسی دانشجو : اعظم سپاهی

هدف

هدف این پروژه این بود که با استفاده از داده‌های مربوط به مشتریان بانک پیش‌بینی کنیم آیا مشتری در نهایت سپرده مدت‌دار را قبول می‌کند یا نه.

توضیح دیتاست:

دیتاست مربوط به کمپین بازاریابی تلفنی یک بانک پرتغالی است و شامل اطلاعاتی مثل سن، شغل، وضعیت تأهل، میزان موجودی، داشتن وام، نوع تماس و غیره می‌باشد.

ستون خروجی (y) دو مقدار دارد:

- yes:

یعنی مشتری سپرده را خریده است.

- no:

یعنی مشتری نخریده است.

مراحل انجام کار:

- 1. داده‌ها با pandas خوانده شدند.
- 2. داده‌های متنی با LabelEncoder به عدد تبدیل شدند.
- 3. داده‌ها به دو بخش آموزش و تست تقسیم شدند.
- 4. داده‌ها با StandardScaler استاندارد شدند.
- 5. مدل GaussianNB آموزش داده شد.
- 6. با داده‌های تست پیش‌بینی انجام شد.
- 7. دقت (Accuracy)، دقت مثبت (Precision)، یادآوری (Recall) و ماتریس ابهام محاسبه شدند.

نتایج به دست آمده از یک انجام کد و آموزش:

معیار	مقدار تقریبی
Accuracy	0.87
Precision	0.79
Recall	0.73

توضیح کوتاه الگوریتم

الگوریتم Naive Bayes بر پایه‌ی قانون بیز است و فرض می‌کند ویژگی‌ها از هم مستقل هستند. این مدل ساده و سریع است و برای مسائل دسته‌بندی دودویی (مثل بله/خیر) خوب عمل می‌کند.

نتیجه گیری:

مدل Naive Bayes با دقت حدود ۸۷٪ توانست تشخیص دهد که کدام مشتری احتمال بیشتری دارد سپرده را قبول کند برای بهبود نتایج می‌توان ویژگی‌های بهتر انتخاب کرد یا از مدل‌های دیگر مثل Random Forest یا Logistic Regression استفاده کرد.