Medical Diagnosis with Naive Bayes

SELVARAJ.X(2577127)

13-OCT-2023

Introduction

- ❖ Medical diagnosis is a critical domain where accurate and timely assessments can significantly impact patient outcomes.
- ❖ In this case study, we explore the application of the Naive Bayes algorithm to develop a diagnostic system for a medical research institute.
- ❖ Our objective is to create a classification model that aids in the medical diagnosis process by leveraging a dataset containing various medical test results, patient information, and corresponding diagnoses.

Data Exploration:

❖ Begin with dataset loading using Pandas.
❖ Features include medical test results and patient info, labels are diagnoses.
❖ Distribution of diagnoses: Analyze counts of each diagnosis for class balance.

```
data = pd.read_csv('data/heartdisease.csv')
data.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	targ
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	

```
x = data.iloc[:,:-1].values
y = data.iloc[:,-1:].values
```

Data Preprocessing

- ❖ Handle missing values: Impute or remove data points with missing values.
- ❖ Normalize or scale features if required for consistent magnitude.
- ❖ Encode categorical variables into numerical format.
- ❖ Calculate prior probabilities: P(Condition) and P(No Condition) based on class distribution.

Implementing Naive Bayes

- Implement Gaussian,Multinomial and Bernoulli Naive Bayes using scikit-learn.
- Split dataset into training and testing sets to evaluate the models.
- Feature Engineering:
  - Convert test results and patient info into features: Use one-hot encoding or numerical representation.
  - Importance of feature selection: Reducing dimensionality can improve model performance and interpretability.

```python
# Gaussian Naive Baye

gauss_nb = GaussianNB()
gauss_nb.fit(xtrain,ytrain)

ypred = gauss_nb.predict(xtest)
```

```python
# Multinomial Naive Baye

from sklearn.naive_bayes import MultinomialNB

multi_nb = MultinomialNB()
multi_nb.fit(xtrain,ytrain)

ypred = multi_nb.predict(xtest)
```

```python
#Bernoulli  Naive Baye

from sklearn.naive_bayes import BernoulliNB

ber_nb = BernoulliNB()
ber_nb.fit(xtrain,ytrain)

ypred = ber_nb.predict(xtest)
```

Model Training and Model Evaluation

- ❖ Train Naive Bayes model on feature-engineered dataset.
- ❖ Probability estimation in Naive Bayes: Use conditional probabilities to calculate likelihood and apply Bayes' theorem.
- ❖ Assess model using evaluation metrics: Accuracy, precision, recall, F1-score.
- ❖ Interpret results: Discuss the model's ability to correctly classify medical conditions.

```
Accuracy score :          0.8947368421052632

Confusion Martix :
 [[32  4]
 [ 4 36]]

classification Report :
              precision    recall  f1-score   support

           0       0.89      0.89      0.89        36
           1       0.90      0.90      0.90        40

    accuracy                           0.89        76
   macro avg       0.89      0.89      0.89        76
weighted avg       0.89      0.89      0.89        76
```

★ Gaussian Naive Bayes has Higher accuracy than other naive bayes models

Laplace Smoothing

```
Accuracy score :        0.8157894736842105

Confusion Martix :
 [[27  9]
 [ 5 35]]

classification Report :
              precision    recall  f1-score   support

           0       0.84      0.75      0.79        36
           1       0.80      0.88      0.83        40

    accuracy                           0.82        76
   macro avg       0.82      0.81      0.81        76
weighted avg       0.82      0.82      0.81        76
```

★ After perform laplace smoothing on Multinomial Naive model has no effect

Real-World Application

- Importance of accurate medical diagnosis: Crucial for patient treatment and outcomes.

- Practical implications of Naive Bayes: Faster, interpretable, and useful for initial screening or decision support in healthcare.

Model Limitations

- ❖ Potential limitations of Naive Bayes: Assumes independence, may not capture complex relationships.

- ❖ Scenarios where it may not perform well: Complex, high-dimensional data with strong dependencies.

Conclusion

- The application of the Naive Bayes algorithm in the context of medical diagnosis offers a valuable tool for healthcare and research.
- Through this case study, we have seen the importance of thorough data exploration, preprocessing, and feature engineering to prepare medical data for modeling.
- The use of Multinomial and Bernoulli Naive Bayes variants, as implemented in scikit-learn, allows for effective classification.