

Packaging the ML Model of Classification

Problem Statement

- Company wants to automate the loan eligibility process based on customer detail provided while filling online application form.
- It is a classification problem where we have to predict whether a loan would be approved or not.

Data

The data corresponds to a set of financial requests associated with individuals.

Variables	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income
CoapplicantIncome	Coapplicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	Loan approved (Y/N)

Source: Kaggle

Running Locally

Add PYTHONPATH variable for `~/ .bash_profile` for MacOS ``export

PYTHONPATH="/Users/nachiketh/Desktop/author-repo/Complete-MLOps-BootCamp/Packaging-ML-Model/packaging-ml-model:\$PYTHONPATH"

```
## Virtual Environment
Install virtualenv

```python
python3 -m pip install virtualenv
```

Check version

```
virtualenv --version
```

Create virtual environment

```
virtualenv ml_package
```

Activate virtual environment

For Linux/Mac

```
source ml_package/bin/activate
```

For Windows

```
ml_package\Scripts\activate
```

Deactivate virtual environment

```
deactivate
```

## Directory structure

---

```
prediction_model

■■■ MANIFEST.in
■■■ prediction_model
■ ■■■ config
■ ■ ■■■ config.py
■ ■ ■■■ __init__.py
■ ■■■ datasets
■ ■ ■■■ __init__.py
■ ■ ■■■ test.csv
■ ■ ■■■ train.csv
■ ■■■ __init__.py
■ ■■■ pipeline.py
■ ■■■ predict.py
■ ■■■ processing
■ ■ ■■■ data_handling.py
■ ■ ■■■ __init__.py
■ ■ ■■■ preprocessing.py
■ ■■■ trained_models
■ ■ ■■■ classification.pkl
■ ■ ■■■ __init__.py
■ ■■■ training_pipeline.py
■ ■■■ VERSION
■■■ README.md
■■■ requirements.txt
■■■ setup.py
■■■ tests
 ■■■ pytest.ini
 ■■■ test_prediction.py
```

## Build the Package

---

Goto Project directory and install dependencies `pip install -r requirements.txt`

Create Pickle file after training: `python prediction_model/training_pipeline.py`

Create source distribution and wheel `python setup.py sdist bdist_wheel`

## Installation of Package

---

Go to project directory where `setup.py` file is located

1. To install it in editable or developer mode

```
pip install -e .
```

`.` refers to current directory

`-e` refers to --editable mode

1. Normal installation

```
pip install .
```

`.` refers to current directory

1. Also can be installed from git as well after pushing to github

```
pip install git+https://github.com/manifoldailearning/prediction_model.git
```

## Testing the Package Working

---

1. Remove the PYTHONPATH from environment variables
2. Goto a separate location which is outside of package directory
3. Create a new virtual environment using the commands mentioned above & activate it
4. Before installing, test whether you are able to import the package of `prediction_model` - (you should not be able to do it)
5. Now in the new environment install the package from github `pip install git+https://github.com/manifoldailearning/prediction_model.git`
6. Now try importing the `prediction_model`, you should be able to do it successfully
7. Extras : Run training pipeline using the package, and also conduct the test