Module `src`

Source code of your project

Sub-modules

- src.load_data
- src.metrics
- src.process
- src.train_model

Module `src.load_data`

This is code is used to Load the dataset into memory.

Author: Aravindh P

Functions

Function `load_train_data`

```
def load_train_data(
    paths: dict
) -> tuple[pandas.core.series.Series, pandas.core.series.Series]
```

This functions loads the training data into memory.

Args ——= **`paths`** : This contains the paths of the datasets.

Returns ——= They data is finally returned as dependent features and target variable.

Function `load_val_data`

```
def load_val_data(
    paths: dict
) -> tuple[pandas.core.series.Series, pandas.core.series.Series]
```

This functions loads the validation data into memory.

Args ——= **`paths`** : This contains the paths of the datasets.

Returns ——= They data is finally returned as dependent features and target variable.

Module `src.metrics`

This is code is used to calculate the metrics for the given model predictions.

Author: Aravindh P

Functions

Function `evaluate_model`

```
def evaluate_model(
    y_true,
    y_pred
) -> dict
```

This function is used to calculate the classification model metrics.

Args ——= **`y_true`** : This is the actual values.

`y_pred` This is the predicted values.

Returns ——= It finally calculates all the metrics and returns in a dictionary.

Module `src.process`

This is code is used to process the dataset.

Author: Aravindh P

Functions

Function `process_data`

```
def process_data(
    config: omegaconf.dictconfig.DictConfig
)
```

This is the configuration function used process the raw data.

Args ——= **`config`** : This is the YAML config info.

Function `process_text`

```
def process_text(
    path: str,
    col: str,
    batch: int
) -> pandas.core.frame.DataFrame
```

This function is used to process all the sentences in the given data series, by removing stopwords and lemmatization of the words.

Args ——= **`path`** : This is the path to the given data series to be processed.

`col` The column that contains the text data.
`batch` The number of sentences to be processed parallely.

Returns ——= It returns the processed data series data-frame.

Module `src.train_model`

This code is used to train the model. Author: Aravindh P

Functions

Function `create_pipeline`

```
def create_pipeline(
    steps: int,
    class_weights: str = 'balanced',
    jobs=-1
) -> sklearn.pipeline.Pipeline
```

This function is used to create the modeling pipeline.

Args ——= **`steps`** : It is the number of iterations for the model before convergence.

`class_weights` It is the used to balance the target classes if the data is imbalanced.
`jobs` The number of parallel process to run.

Returns ——= The function returns the constructed pipeline.

Function `train_model`

```
def train_model(
    config: omegaconf.dictconfig.DictConfig
)
```

This function is configuration function used to train the model.

Args ——= `config` : This is the YAML config info.
