Bpi deeptech

mardi 24 novembre 2020 10:25

	Model	Features
Main Model	Logistic Regression	'doctor_yesno', 'funding_employees_ratio', 'has_strong_founder', 'has_super_founder', 'stage_age_ratio', 'nb_patents', 'technical_background', 'health_industry', 'semiconductors_industry', 'energy_industry', 'commission_income_streams', 'biotechnology_tags', 'neurology_tags', 'saas_tags', 'fund_investors_type', 'Agoranov_investors_name'
Scores		
Laboratory	Logistic Regression	Zip Code, nb patents, doctors (Yes/No)
Time to market	Logistic Regression	funding_employees_ratio, stage_age_ratio
Technological	Logistic Regression	tags

```
KEPT_TAGS = [
             'technical_background',
              'health_industry',
              'semiconductors_industry',
              'energy_industry',
              'commission_income_streams',
              'biotechnology_tags',
              'neurology_tags',
              'saas_tags',
              'fund_investors_type',
              'Agoranov_investors_name']

TARGET_ZIP = [91, 38, 87, 35, 67]
```

Time to market

stage_age_ratio

```
def get_stage_age_ratio(data):
    data['year'] = pd.DataFrame({'year': data['launch_year']})
    data['year_of_existence'] = data['year'].map(lambda x : substract_date(x))
    data['growth_stage_num'] = growth_stage_num(data)
    data['stage_age_ratio'] = data[['year_of_existence','growth_stage_num']]\
                              .apply(return_ratio,axis=1)
    return data['stage_age_ratio']
```

growth_stage_num

```
def growth_stage_num(data):
    '''
    function that maps the 'growth stage' as an ordinal feature
    '''

    stage_status = data['growth_stage'].map({'mature' : 4,
                                             'late growth' : 3,
                                             'early growth' : 2,
                                             'seed' : 1})
    return stage_status
```

funding_employees_ratio

```
def funding_amounts_employees(data):
    '''
    function that performs the ratio between funding and employees
    '''

    funding = data['total_funding_source']
    employees = data['employees_latest']

    return funding/employees
```