**Lab Regression 1 (8/4/2022)**

ລະຫັດນັກສຶກສາ: 205Q001019

ຊື່ ແລະ ນາມສະກຸນ: ທ ນູຊົ່ວ ເຮີ

ຈົ່ງຕອບຄຳຖາມຕໍ່ໄປນີ້ໃຫ້ສຳເລັດ:

1.  Regression

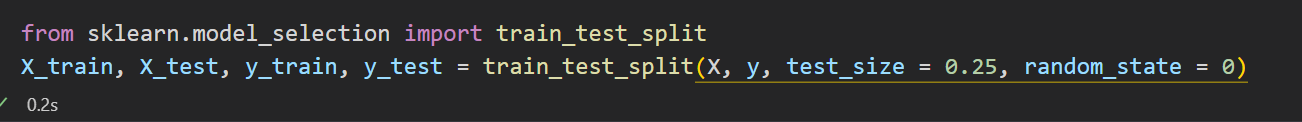
1.1. ຈາກຊຸດຂໍ້ມູນ ( ປີData Set)  Salary\_Data.csv, ຈົ່ງບອກຈຳນວນຖັນ, ແຖວ (shape) ແລະ ເພີ່ມຂໍ້ມູນໃນຕາຕະລາງລຸ່ມນີ້ໃຫ້ສຳເລັດ:

ຈຳນວນຖັນ, ແຖວ (shape)

|  |  |
| --- | --- |
| ຖັນ | 2 |
| ແຖວ | 30 |

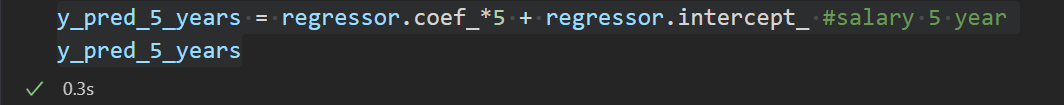
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **YearsExperience** | **Salary** |  |  |
| 0 | 1.1 | 39343.0 |  |  |
| 1 | 1.3 | 46205.0 |  |  |
| 2 | 1.5 | 37731.0 |  |  |
| 3 | 2.0 | 43525.0 |  |  |
| 4 | 2.2 | 39891.0 |  |  |

1.2. ຈາກຊຸດຂໍ້ມູນ (Data Set) Salary\_Data.csv.ຈົ່ງຂຽນຄຳສັ່ງເພື່ອແຍກຊຸດຂໍ້ມູນອອກເປັນສອງພາກສ່ວນຄື: ຊຸດທົດສອບ 75% ແລະ ຊຸດທົດສອບ 25%?



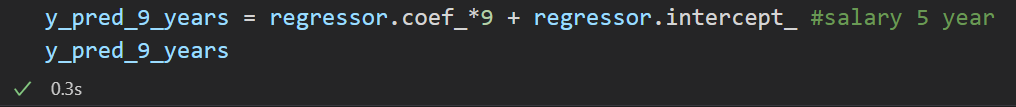
1.3. ຈົ່ງຮຽນຊຸດຂໍ້ມູນຂໍ້ 1.2ດ້ວຍໂມເດວ LinearRegression ເພື່ອຄາດເດົາເງີນເດືອນ (**Salary)** ຂອງພະນັກງານທີມີປະສົບ 5ປີ ແລະ 9ປີ?

* ມີປະສົບ 5ປີ



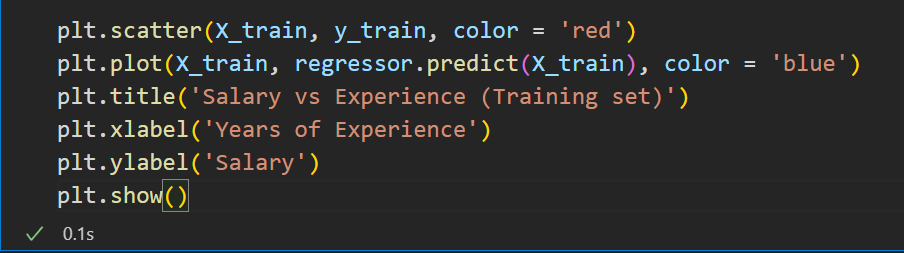
array([73885.24377647])

* ມີປະສົບ 9ປີ



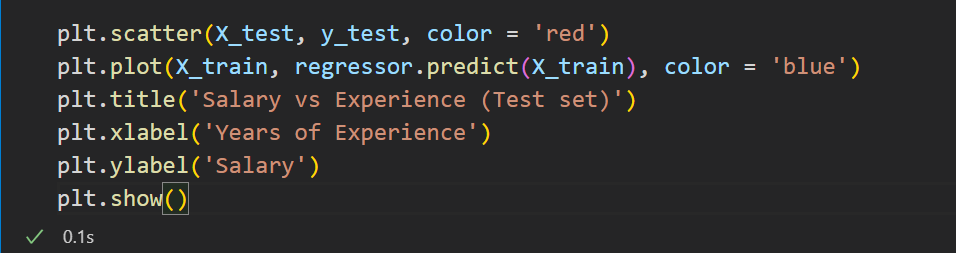
array([111404.08574425])

1.4 ຈົ່ງສະແດງຂໍ້ມູນຊຸດຮຽນ (X\_train, y\_train) ດ້ວຍ scatter ຫຼື seaborn Graph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib





1.5 ຈົ່ງສະແດງຂໍ້ມູນຊຸດທົດສອບ (X\_test, y\_test) ດ້ວຍ scatter ຫຼື seaborn Graph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib





1.6 ຖາມວ່າພະນັກງານສ່ວນຫຼາຍມີເງີນເດືອນຢູ່ຊ່ວງໃດ ແລະ ມີປະສົບການເຮັດວຽກຊ່ວງໃດ

* ພະນັກງານສ່ວນຫຼາຍມີເງິນເດືອນຢູ່ຊ່ວງ 40000-60000
* ມີປະສົບການເຮັດວຽກຊ່ວງ 2-4 ປີ

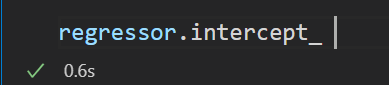
1.7 ຈົ່ງຄິດໄລ່ R squared (ຄວາມຖືກຕ້ອງໃນການຄາດເດົາ)

* regressor.score(X\_train, y\_train)

0.9395413526983522

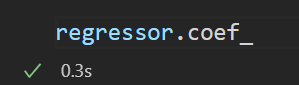
1.8 ເມື່ອ y = ax + b (Simple Linear Regression), ຈົ່ງຄິດໄລ່ b ~~intercept ແລະ a~~- coef

* b -intercept



26986.691316737248

* a-coef



array([9379.71049195])