

PREDICTING SOLAR RADIATION LEVEL WITH ML REGRESSION MODEL



Presented by,
Harsh.

BACKGROUND

SOLAR RADIATION LEVEL PREDICTION-NASA

These datasets are meteorological data from the HI-SEAS weather station from four months (September through December 2016) between Mission IV and Mission V.

For dataset, the fields are:

The units of each dataset are

1. Solar radiation: watts per meter^{^2}
2. Temperature: degrees Fahrenheit
3. Barometric pressure: Hg
4. Wind direction: degrees
5. Wind speed: miles per hour
6. Sunrise/sunset: Hawaii time



OBJECTIVE

To predict the Solar radiation levels given the measurements temperature, barometric pressure, wind direction & wind speed.



DATA DESCRIPTION

RangeIndex: 32686 entries, 0 to 32685

Data columns (total 6 columns):

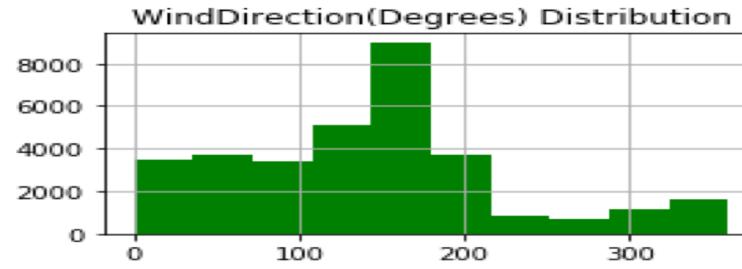
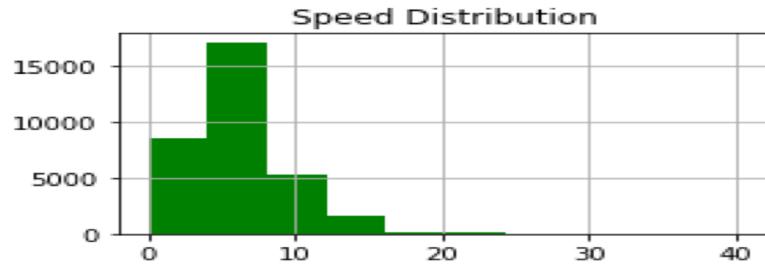
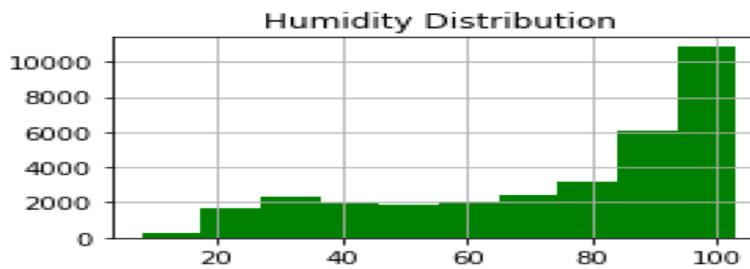
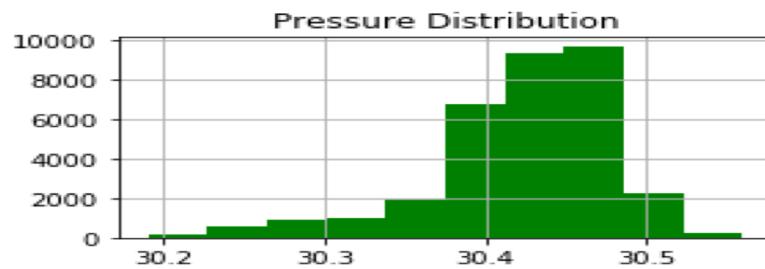
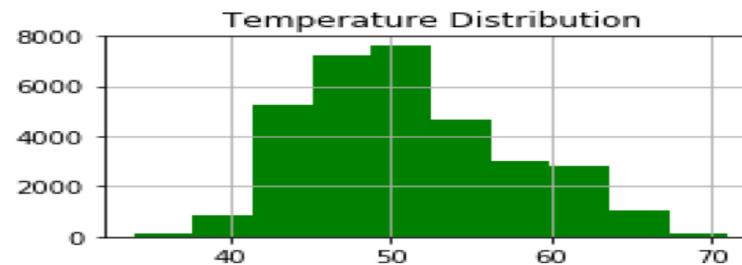
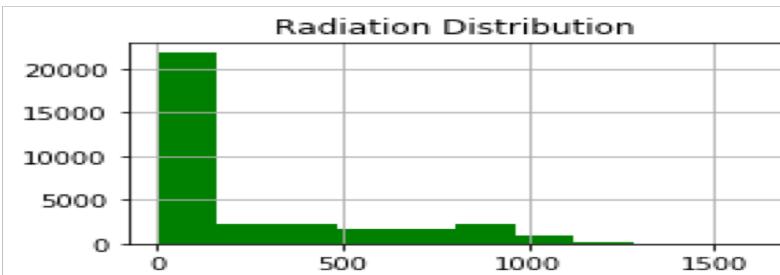
1. Radiation 32686 non-null float64
2. Temperature 32686 non-null int64
3. Pressure 32686 non-null float64
4. Speed 32686 non-null float64
5. WindDirection(Degrees) 32686 non-null float64

#CHECKING FOR MISSING VALUES

Missing value in the dataset					
Radiation -	Temperature -	Pressure -	Humidity -	Speed -	WindDirection(Degrees) -
100	100	100	100	100	100
100	100	100	100	100	100
100	100	100	100	100	100

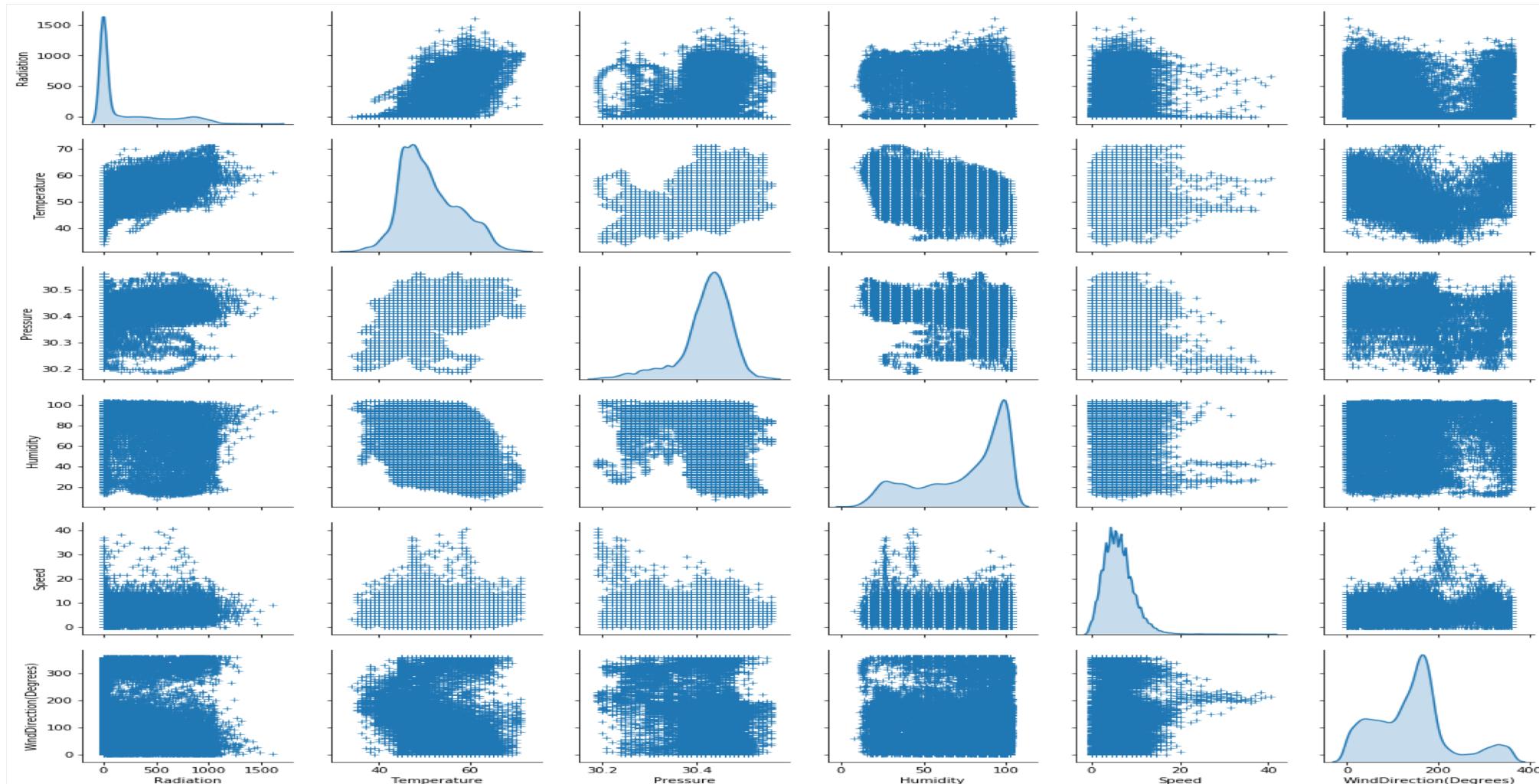


#DESCRIBING FEATURE DISTRIBUTION



We can easily see that Temperature and wind direction is almost normally distributed while radiation and humidity is positive and negative skewed respectively.

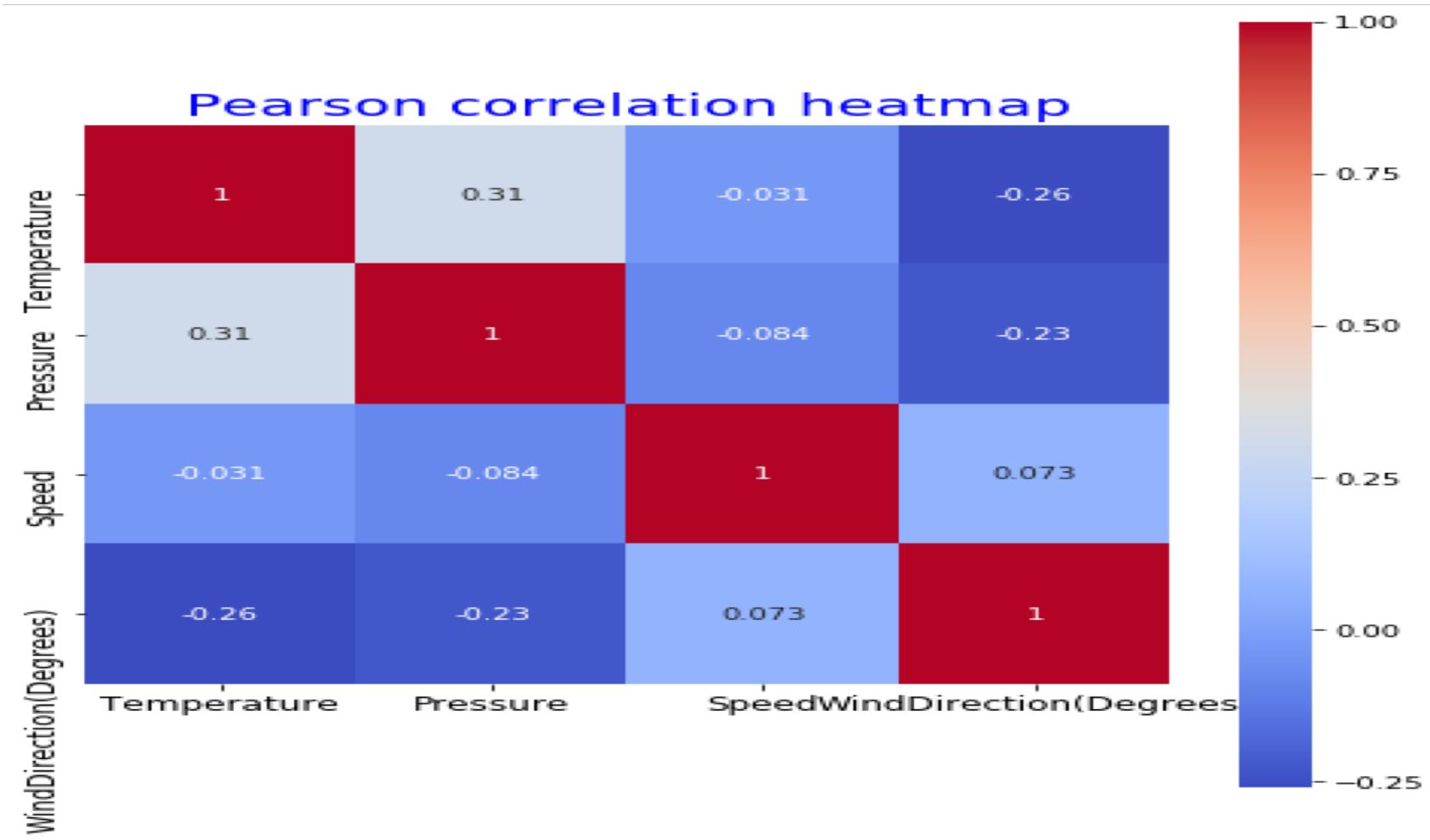
FOR FURTHER VISUALIZE ANY RELATIONSHIPS BETWEEN THE FEATURES,



It can easily seen from above pair plot that temperature and radiation have some linear relationship.



PEARSON CORRELATION HEATMAP (MATRIX) WAS PLOTTED FOR CORRELATION.



It is clearly visible from above picture that there is no correlation between all the independent variables.

DEPENDENT & INDEPENDENT VARIABLES

Dependent variable -Radiation.

Independent variables- Temperature,Pressure,Speed, wind direction

#K-FOLD CROSS VALIDATION

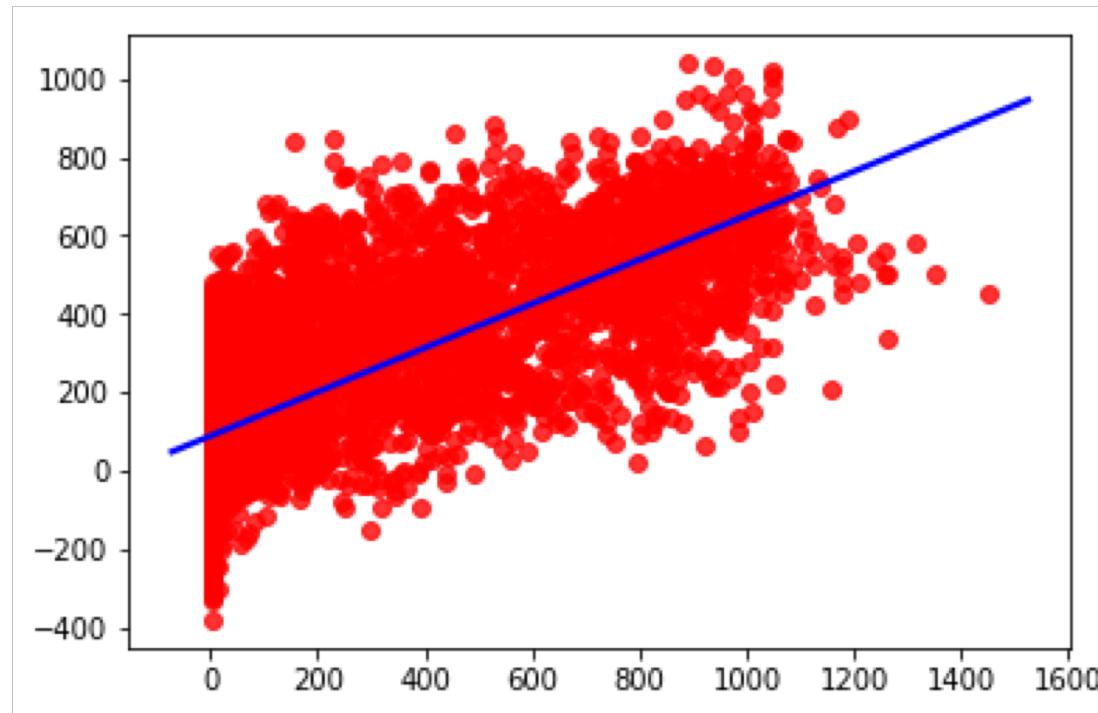
- For k fold validation we have split data into 80:20 train and test and then we have trained our model with linear regression model.
- K- fold validation has been performed with 10 folds for which average score is 0.56

R2,ADJUSTED R2 & P-VALUE FOR EACH INDICATOR

OLS linear square model has also been executed for R2, adjusted R2 and p values for each independent variables and they found to be very significant.

OLS Regression Results						
Dep. Variable:	Radiation	R-squared:	0.566			
Model:	OLS	Adj. R-squared:	0.565			
Method:	Least Squares	F-statistic:	1.063e+04			
Date:	Tue, 16 Oct 2018	Prob (F-statistic):	0.00			
Time:	05:28:22	Log-Likelihood:	-2.2088e+05			
No. Observations:	32686	AIC:	4.418e+05			
Df Residuals:	32681	BIC:	4.418e+05			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2.023e+04	682.829	29.625	0.000	1.89e+04	2.16e+04
Temperature	38.6648	0.200	193.681	0.000	38.274	39.056
Pressure	-723.5344	22.516	-32.134	0.000	-767.667	-679.402
Speed	8.3323	0.332	25.121	0.000	7.682	8.982
WindDirection(Degrees)	-0.2615	0.015	-17.966	0.000	-0.290	-0.233
Omnibus:	2370.776	Durbin-Watson:	0.220			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3286.673			
Skew:	0.624	Prob(JB):	0.00			
Kurtosis:	3.925	Cond. No.	1.03e+05			

#VALIDATING MODEL BETWEEN PREDICTED RADIATION AND ACTUAL RADIATION.



Above graph shows linear relationship within predicted and test y variables.

CONCLUSION.

- 68 % change in radiation can be explained by 'Temperature','Pressure', 'Speed','WindDirection(Degrees)' variables.
- As per validation of model there is a linear relationship between predicted radiation and actual radiation which shows good fitting up to certain extent and it can predict the radiation level by inserting the independent variables values.

Thank you!

