Problem Statement

Overview

Data scientist is the sexiest job in the world. How many times have you heard that? Analytics India Annual Salary Study which aims to understand a wide range of trends data science says that the median analytics salary in India for the year 2017 is INR 12.7 Lakhs across all experience level and skill sets. So given the job description and other key information can you predict the range of salary of the job posting? What kind of factors influence the salary of a data scientist? The study also says that in the world of analytics, Mumbai is the highest paymaster at almost 13.3 Lakhs per annum, followed by Bengaluru at 12.5 Lakhs. The industry of the data scientist can also influence the salary. Telecom industry pays the highest median salaries to its analytics professionals at 18.6 Lakhs. What are you waiting for, solve the problem by predicting how much a data scientist or analytics professional will be paid by analysing the data given. Bonus Tip: You can analyse the data and get key insights for your career as well. The best data scientists and machine learning engineers will be given awesome prizes at the end of the hackathon. Share this hackathon with a colleague who may be interested in mining the dataset for insights and make great predictions. Data The dataset is based on salary and job postings in India across the internet. The train and the test data consists of attributes mentioned below. The rows of the train dataset has a rich amount of information regarding the job posting such as name of the designation and key skills required for the job. The training data and test data consist of 19802 samples and of 6601 samples each. This is a dataset which has been collected over some time to gather relevant analytics jobs posting over the years. Features Name of the company (Encoded) Years of experience Job description Job designation Job Type Key skills Location Salary in Rupees Lakhs(To be predicted) Problem Statement Based on the given attributes and salary information, build a robust machine learning model that predicts the salary range of the salary post.

[Download data from here](#)

You're expected to write proper commented code with explaining the entire pipeline for machine learning project: starting from data cleaning, selecting model to Visualization.