Using two regression models to predict the salary and evaluate their strengths

The aim of my project work is to use polynomial regression and support vector regression models. Precisely speaking, the objective is to use the polynomial regression and support vector model to predict what would be the salary of the person at a specific position. Both polynomial regression and support vector model have origins in the field of supervised learning.

As far as polynomial regression is concerned, briefly, the idea is to fit a polynomial equation on the data with a relationship between the independent variable and dependent variable. On the other hand, as far as support vector regression is concerned, in brief, the goal is to find the linear and nonlinear relationship and to fit the the values in the model.

The aim here is to test these two models and evaluate their strengths and weaknesses. First, I tried to look a perfect dataset and approached the Statistics Centre here in Finland but I have not yet received feedback if there is a dataset available. If I am not capable to get the dataset from Finland I will use the following the dataset downloaded from Kaggle:

https://www.kaggle.com/kaggle/sf-salaries?select=Salaries.csv

The dataset includes 148 655 data rows or so called input features with information on id, employee name, job title, base pay, overtime pay, other pay, benefits, total pay, total pay benefits, year, notes, agency, and status. This data contains the names, job title, and compensation for San Francisco city employees from 2011 to 2014. The labels have output information about these input features. Data points contains, for instance, the values of salary and the job title. When I have trained two selected models the goal is return the predicted label for the certain job title.