DATA MINING COURSE PROJECT

SUBMITTED TO : DR. NAGAMMA PATIL
SUBMITTED BY :
APOORVA CHANDRA(11IT09)
DILIP MALLYA(11IT19)
JNVK CHAITANYA(11IT32)
R. KAUSHIK(11IT65)

INTRODUCTION

- Forecasting is necessary in many fields – be it weather , military application, sports and so on

- Forecasting finds a special place in business and finance

INTRODUCTION

- In current times, social media is a very active hub of interaction – many businessmen use it to find potential markets and customers

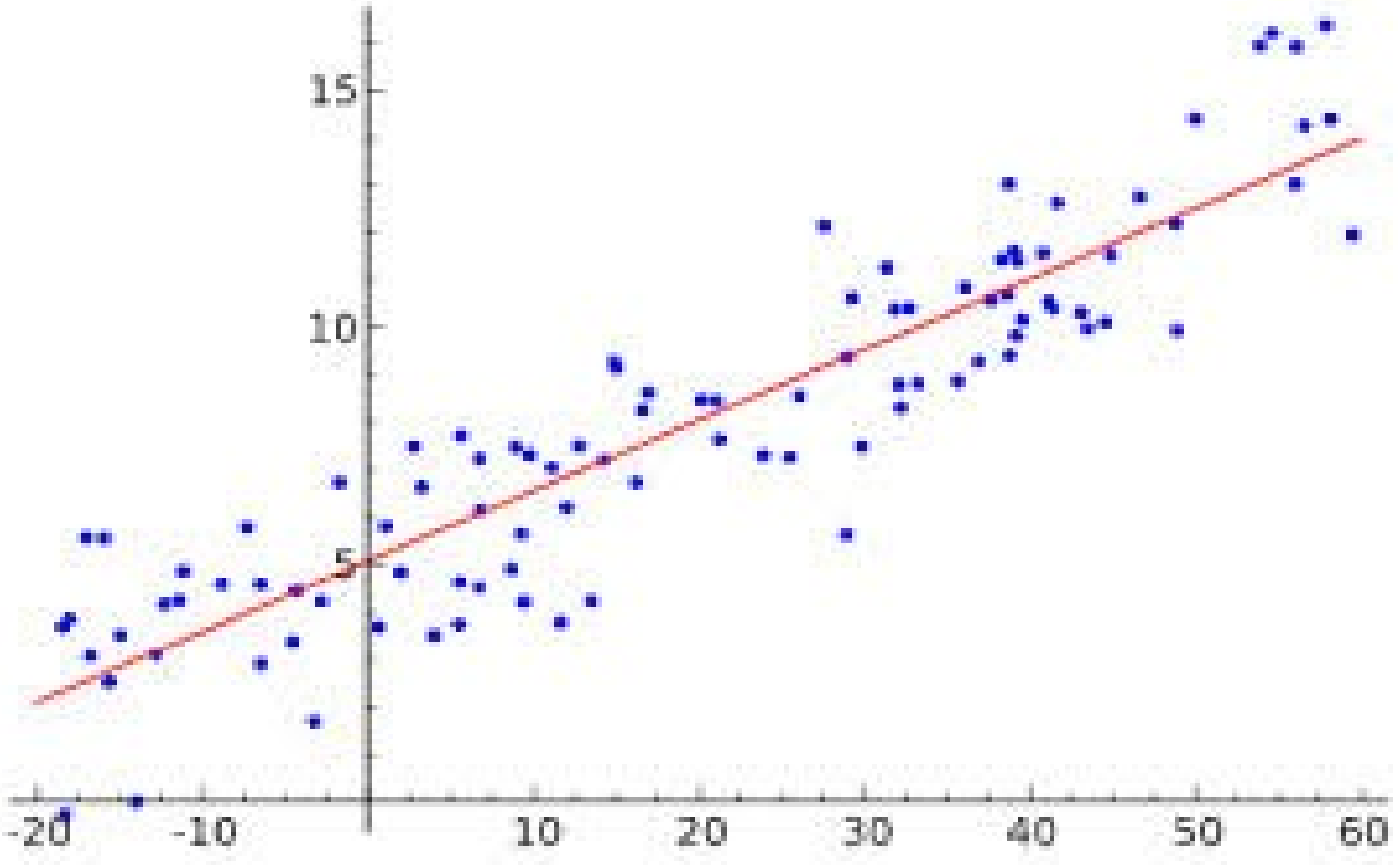
- Forecasting employed in social media can offer many benefits

POPULAR TECHNIQUES IN FORECASTING

- Linear regression
- ARMA
- VAR
- ARIMA

LINEAR REGRESSION

- Predicting a most probable line through a set of points such that sum of square of errors in minimum



FORMULAE

$$y = \alpha + \beta x_i + \varepsilon_i$$

$$E(\varepsilon_i) = 0, VAR(\varepsilon_i) = \sigma^2$$

FORMULAE

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})/n}{(x_i - \overline{x})^2}$$

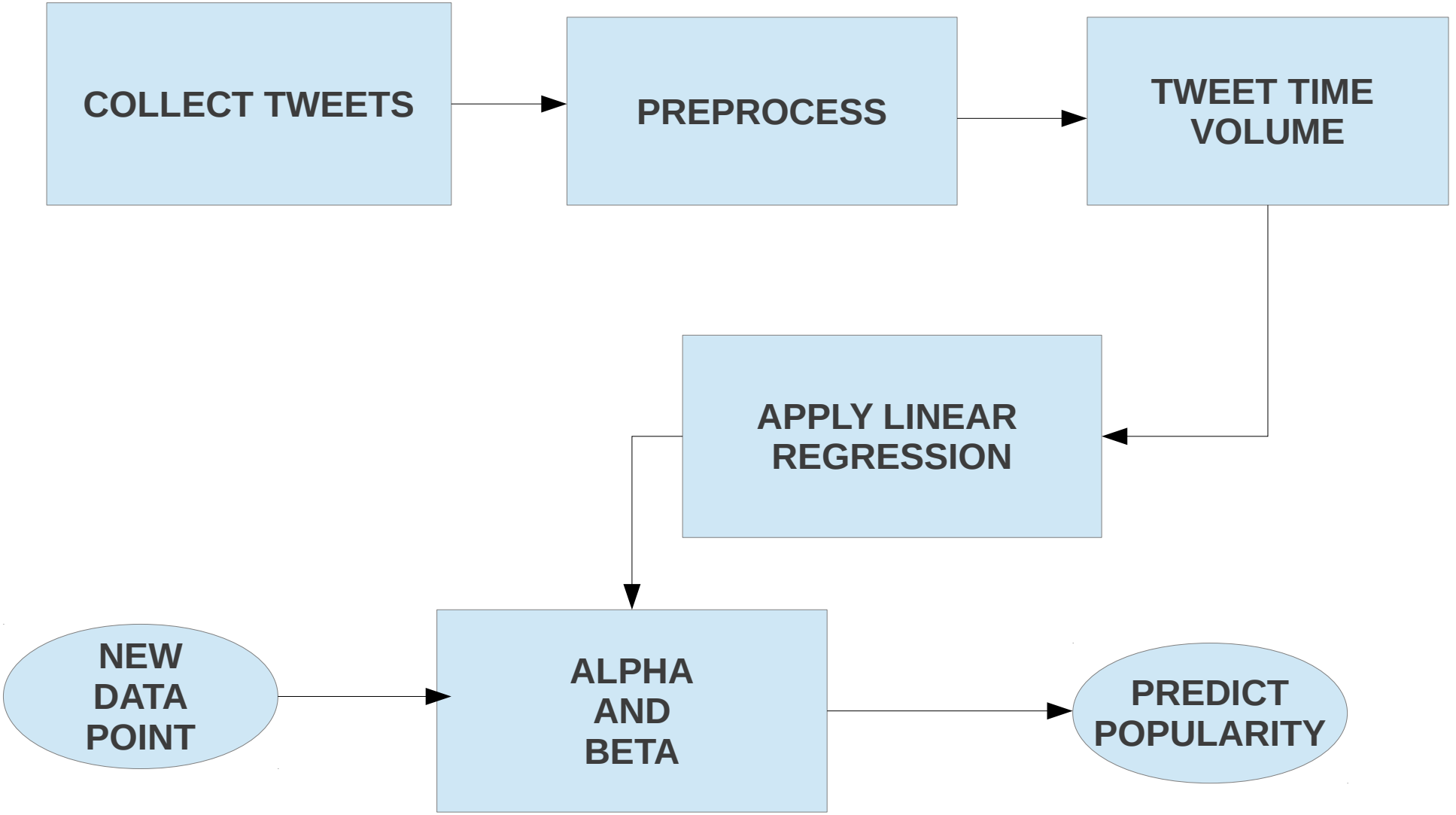
FORMULAE

$$\hat{\alpha} = \overline{y} - \beta\overline{x}$$

PAPER IMPLEMENTED

- Predicting Future Popularity Trend of Events in Microblogging Platforms - Manish Gupta et al

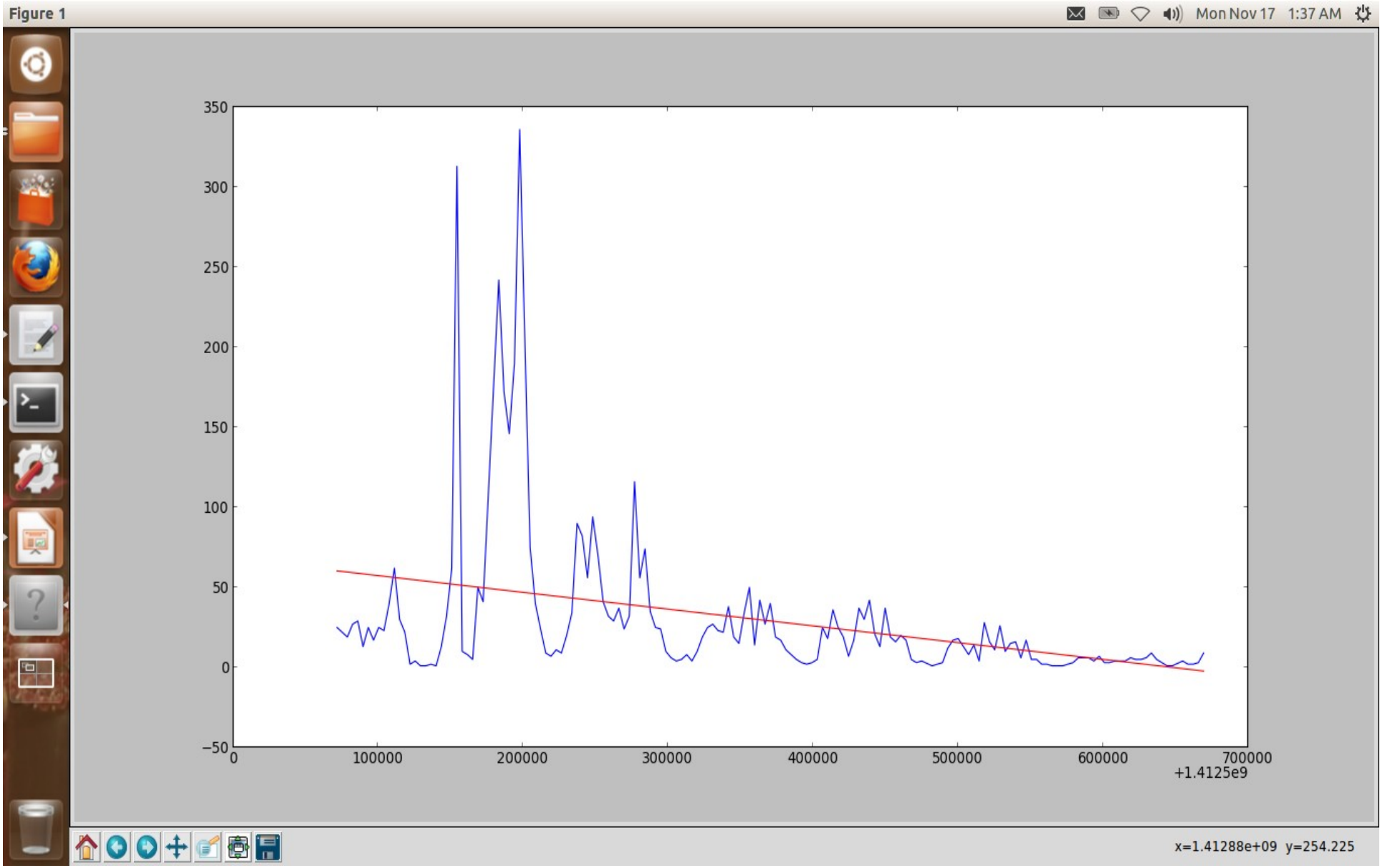
METHOD


COLLECT TWEETS
PREPROCESS
TWEET TIME VOLUME
APPLY LINEAR REGRESSION
NEW DATA POINT
ALPHA AND BETA
PREDICT POPULARITY

DATASETS USED

- Jayalalitha dataset
- Hudhud dataset
- Ebola dataset

RESULTS – jayalalitha dataset



RESULTS

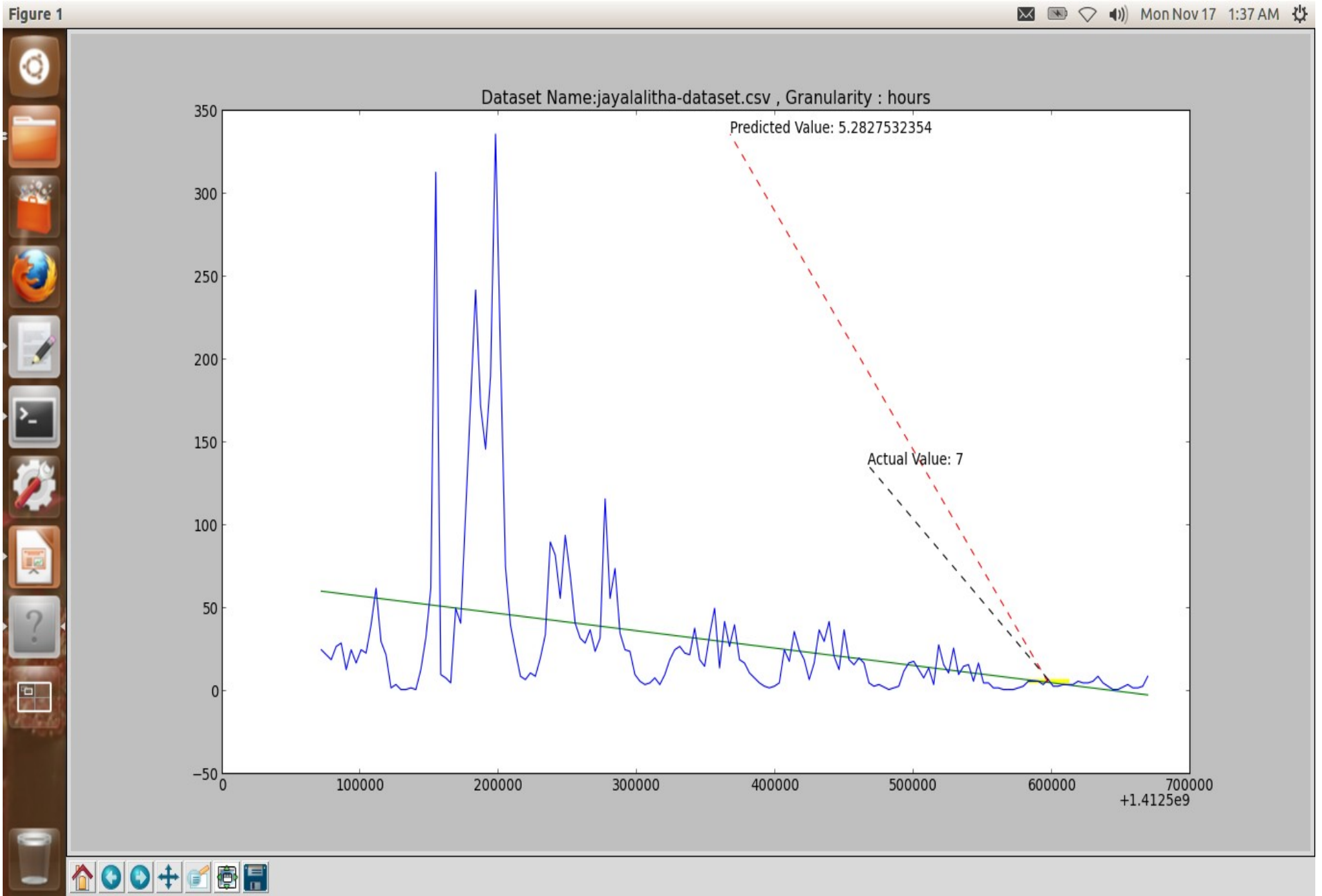
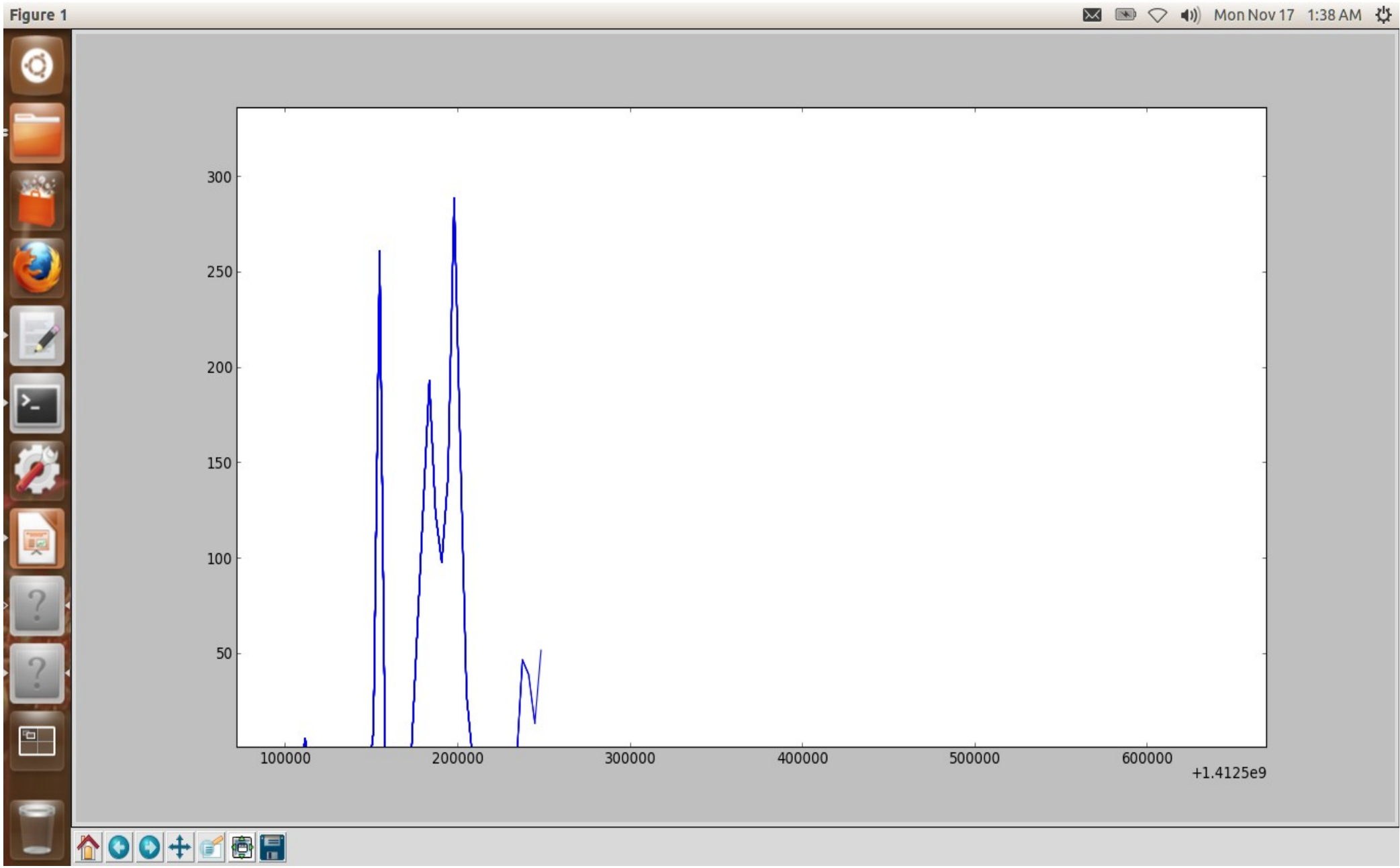


Figure 1
Mon Nov 17 1:37 AM
Dataset Name:jayalalitha-dataset.csv , Granularity : hours
Predicted Value: 5.2827532354
Actual Value: 7
350
300
250
200
150
100
50
0
−50
0
100000
200000
300000
400000
500000
600000
700000
+1.4125e9

RESULTS - ERRORBAR



CONCLUSION

- Linear regression yields better results on datasets processed with:

  Lesser granularity
- This is because of :

  1) more data points or tweets

  2) more distributed

QUESTIONS