**Cab Cancellation Prediction**

**-Manish Kumar[1PE16IS053]**

**-Manasa Hegde[1PE16IS052]**

**Problem Statement**

The business problem tackled here is trying to improve customer service for [YourCabs.com](http://www.yourcabs.com/), a cab company in Bangalore. The problem of interest is booking cancellations by the company due to unavailability of a car. The challenge is that cancellations can occur very close to the trip start time, thereby causing passengers inconvenience.

**Goal**

The goal is to create a predictive model for classifying new bookings as to whether they will eventually get cancelled due to car unavailability. This is a classification task that includes misclassification costs.

**File descriptions**

Kaggle\_YourCabs\_training.csv - the training set (over 43,000 bookings). Includes the output Car\_Cancellation and the misclassification costs in Cost\_of\_error.

Kaggle\_YourCabs\_score.csv - the data set to be classified. Includes 10,000 bookings and no output column.

Kaggle\_YourCabs\_sample.csv - a sample submission file in the correct format. Your entry should include the id column from this file and a Car\_Cancelled column with 0,1 values.

**Data fields**

id - booking ID

user\_id - the ID of the customer (based on mobile number)

vehicle\_model\_id - vehicle model type.

package\_id - type of package (1=4hrs & 40kms, 2=8hrs & 80kms, 3=6hrs & 60kms, 4= 10hrs & 100kms, 5=5hrs & 50kms, 6=3hrs & 30kms, 7=12hrs & 120kms)

travel\_type\_id - type of travel (1=long distance, 2= point to point, 3= hourly rental).

from\_area\_id - unique identifier of area. Applicable only for point-to-point travel and packages

to\_area\_id - unique identifier of area. Applicable only for point-to-point travel

from\_city\_id - unique identifier of city

to\_city\_id - unique identifier of city (only for intercity)

from\_date - time stamp of requested trip start

to\_date - time stamp of trip end

online\_booking - if booking was done on desktop website

mobile\_site\_booking - if booking was done on mobile website

booking\_created - time stamp of booking

from\_lat - latitude of from area

from\_long - longitude of from area

to\_lat - latitude of to area

to\_long - longitude of to area

Car\_Cancellation (available only in training data) - whether the booking was cancelled (1) or not (0) due to unavailability of a car.

Cost\_of\_error (available only in training data) - the cost incurred if the booking is misclassified. For an un-cancelled booking, the cost of misclassificaiton is 1. For a cancelled booking, the cost is a function of the cancellation time relative to the trip start time (see Evaluation Page).

**Techniques used**

KNN and SVM is applied.