**Machine Learning Problem**

**Mess Food Saver**

Mess Wastage is a very common problem that affects almost every college. We are developing a ML based solution, which will try to solve the problem by predicting the number of students going to dine in the Mess. We have been provided one-year historical data corresponding to following features stated below:

1.     Date – DD/MM/YYYY

2.     Month   - MM

3.     Day of the Week- Derivable from Date

4.     Type\_of\_Meal

5.     No\_of\_Students\_in\_hostel

6.     No\_of\_actual\_diners

7.     Special\_Occasion – Pertaining to Exams, Holidays, Major Events etc

Let’s say a mess manager has to decide the quantity of food to be prepared for the students in a shift. It is reasonable to assume that Mess Manager will want predictions for the next few shifts (lets say 5) at a particular point of time, and the range of prediction can be adjusted as per the needs. We have to assume that our model has a real time access to data pertaining to Number of Students in hostel.  Let’s assume that ideal time for food preparation is around 2-3 hours. So in a way Mess Manager has to decide on the Food Quantity to be prepared 2-3 hours back of actual time the food is going to be served.  The issue here is, if we are going to predict the number of diners for lunch, then assuming students will be having classes, we have to decide on certain cutoff time to take in the value of no of students in hostel,e.g. if lunch is going to be served at 1:00 PM then students wont be available at 10:00 AM in hostel and real  time data corresponding to number of students wont be of any use. So in this case the generalised feature representation for the model to use should be changed a bit and should include the number of students available and dined in last four shifts as separate features.

So now the feature representation looks like following:

Date, Month, Day of Week, Type of Meal, No of Students in Hostel(Real Time), Special Occasion, **No of Student in Hostel in Shift-1, Number of Students dined in PresentShiftminus1, No of Student in Hostel in PresentShiftminus2, Number of Students dined in PresentShiftminus2, No of Student in Hostel in PresentShiftminus3, Number of Students dined in PresentShiftminus 3, No of Student in Hostel in PresentShiftminus4, Number of Students dined in PresentShiftminus4**

The additional time series based features are in bold and that will take care of real time mismatch plus the recent trends.

We assume that Mess Manager is always informed of some unscheduled events and occasions that might be happening in college which might affect the number of people in the hostel.  These scheduled/unscheduled events might be following:

**Scheduled:**  Semester Break, Long Holidays, Short Holidays, Exams, Seminars, Fests etc

Long Holidays means when a particular holiday is a part of weekends etc which makes people to go on trips/home etc

Short Holidays: Other than Long holidays and weekends

Exams: During exams there might be a mismatch in people in hostel and people coming to dine as lot of people like to study and may skip food. We will also include a **feature to capture the special event (in case of dinners or may be others too) just after the day of exams when people most of the times eat outside**.

College Fests/Seminars: In case of fests/seminars/functions which may or may not extend upto three or more days, the things become tricky as the features which we have included till now would not be able to resonate with the number of people in hostel and in dining. So we will include another feature which will represent the number of students in hostel before the event started and number of people dined before the event started.

The above are scheduled events and we expect Mess manager to know these events and enter the same in the portal for getting the right prediction for next Shift. There might be other unscheduled events like closing of food shops outside college etc which will make all students to dine in the mess or some protest etc which might decrease people both in hostel and in mess. So in these situations a Mess Manager can classify the next shift to be a part of the category described in the Scheduled section based on the effect he perceives due to that unscheduled event. If there are wide variations in the model prediction and Actual number of people dining or If the mess manager perceives a very different effect, we have also provided a manual correction factor for adjusting the predictions coming out of model and this factor would be absolutely at the discretion of the Mess Manager.

With the above arguments following is the updated feature representation:

Date, Month, Day of Week, Type of Meal, No of Students in Hostel(Real Time), Special Occasion, **No of Student in Hostel in Shift-1, Number of Students dined in PresentShiftminus1, No of Student in Hostel in PresentShiftminus2, Number of Students dined in PresentShiftminus2, No of Student in Hostel in PresentShiftminus3, Number of Students dined in PresentShiftminus 3, No of Student in Hostel in PresentShiftminus4, Number of Students dined in PresentShiftminus4, Average Number of Students dined in 3 shifts before event, Average Number of Students present in hostel in 3 shifts before event**

In case the shift falls in regular event, the last two features will just have last three shift average values.

We also believe that some other additional features will help us in improving the model to a greater extent. This may be a work of future data collection process which usually forms a part of entire Machine Learning project cycle.

**Draft of Additional Features that can improve Model**

1.     Mess\_No

2.     New\_Menu  -- Hypothesis being that new menu may make people to come in large numbers)

3.     Food\_Prepared\_By1 – That can be an important feature if competence is varied across Chefs

4.     Food\_Prepared\_By2

5.     Food\_Prepared\_By3

6.     Meal\_1

7.     Meal\_2

8.     Meal\_3

9.     Meal\_1\_ingredient\_1

10.  Meal\_1\_ingredient\_2

11.  Meal\_1\_ingredient\_3

12.  Meal\_2\_ingredient\_1

13.  Meal\_2\_ingredient\_2

14.  Meal\_2\_ingredient\_3

15.  Meal\_3\_ingredient\_1

16.  Meal\_3\_ingredient\_2

17.  Meal\_3\_ingredient\_3

We believe knowing the count of diners to be present in the Mess is not enough to solve the problem of Mess Food Wastage. Knowing the exact reasons on why the food was wasted will be helpful in making the correct decisions pertaining to saving food and also will help in making necessary machine learning model refinements. That’s how the both the given and draft of additional features will be helpful in determining the reasons behind what may be the cause of the food wastage. This can help in redesigning the menu. We ought to take a Analytical approach to find the reasons behind the wastage and get it proven statistical way.

**Machine Learning**

We will fit following models to get the larger perspective:

1.     Linear Regression Model (To get the interpretability aspects, coefficient, p-values etc)

2.     Boosting Models (To get Feature Importances etc)

3.     Neural Net Models (To improve accuracy etc)

Now shortage of food will be dependent on the fact on how much food is consumed and how much food was prepared initially (which is dependent on how many people our model predicted for). If the food wastage data is collected wrt menu list, then we can get the exact statistics on how the different menu items have performed in terms of wastage. Because different people may like the same item on menu and dislike the other same item, this will lead to wastage in menu items that wont be correlated wrt the mismatch in number of students actually dined and what our model predicted. To tackle with this, adjustment factors with respect to items will be there which will change per student consumption so as to optimize the quantity of food prepared.