**CODEBOOK**

“Getting and Cleaning Data” Course Project Code Book

**Initial data for research**

The data is taken from UCI HAR Dataset. This dataset provides the following variables for each activity:

subject - ID of participant activity - ID of activity type Mean and standart deviation for the following features (other values are presented in initial dataset, but for this reasearch only these parameters were used)

tBodyAcc-XYZ;

tGravityAcc-XYZ;

tBodyAccJerk-XYZ;

tBodyGyro-XYZ;

tBodyGyroJerk-XYZ;

tBodyAccMag;

tGravityAccMag ;

tBodyAccJerkMag;

tBodyGyroMag;

tBodyGyroJerkMag;

fBodyAcc-XYZ ;

fBodyAccJerk-XYZ;

fBodyGyro-XYZ;

fBodyAccMag fBodyAccJerkMag;

fBodyGyroMag fBodyGyroJerkMag;

The features come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix ‘t’ to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz. 1

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag). Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the ‘f’ to indicate frequency domain signals). These signals were used to estimate variables of the feature vector for each pattern: ‘-XYZ’ is used to denote 3-axial signals in the X, Y and Z directions.

**CodeBook**

The following data transformations were conducted to form a tidy dataset: Added a new feature activitylabel - factor variable for activities with the following levels: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING. Tidy dataset was build as a mean values of features grouped by activitylabel and subject - for each subject and activity type determined mean values over all activities of that type.