This document summarizes the preprocessing that has been done per dataset.

For Dataset Assistment skill builder 2009 (the wrong datas):

data = data[['user\_id','skill\_id','skill\_name','correct']]

# Replace multiply appears

data.loc[data['skill\_id']==163,'skill\_id']=85

#Repair wrong words in skill names

data.loc[data['skill\_name']=='Effect of Changing Dimensions of a Shape Prportionally','skill\_name']='Effect of Changing Dimensions of a Shape Proportionally'

data.loc[data['skill\_name']=='Parts of a Polyomial, Terms, Coefficient, Monomial, Exponent, Variable','skill\_name']='Parts of a Polynomial Terms Coefficient Monomial Exponent Variable'

data.loc[data['skill\_name']=='Order of Operations +,-,/,\* () positive reals','skill\_name']='Order of Operations addition subtraction division multiplication parentheses positive reals'

data.loc[data['skill\_name']=='D.4.8-understanding-concept-of-probabilities','skill\_name']='Understanding concept of probabilities'

data.loc[data['skill\_name']=='Angles - Obtuse, Acute, and Right','skill\_name']='Angles Obtuse Acute and Right'

For Dataset Assistment skill builder 2009 corrected:

data = data[['user\_id','skill\_id','skill\_name','correct']]

# Replace multiply appears

data.loc[data['skill\_id']=='163','skill\_id']='85'

data.loc[data['skill\_id']=='1\_13','skill\_id']='1'

data.loc[data['skill\_id']=='1\_15','skill\_id']='1'

data.loc[data['skill\_id']=='10\_104','skill\_id']='10'

data.loc[data['skill\_id']=='10\_12','skill\_id']='10'

data.loc[data['skill\_id']=='10\_13','skill\_id']='10'

data.loc[data['skill\_id']=='10\_14','skill\_id']='10'

data.loc[data['skill\_id']=='10\_15','skill\_id']='10'

data.loc[data['skill\_id']=='10\_18','skill\_id']='10'

data.loc[data['skill\_id']=='10\_58','skill\_id']='10'

data.loc[data['skill\_id']=='10\_64','skill\_id']='10'

data.loc[data['skill\_id']=='11\_70','skill\_id']='11'

data.loc[data['skill\_id']=='16\_17','skill\_id']='16'

data.loc[data['skill\_id']=='173\_190\_193\_221','skill\_id']='170'

data.loc[data['skill\_id']=='2\_37\_48','skill\_id']='2'

data.loc[data['skill\_id']=='2\_37\_70','skill\_id']='2'

data.loc[data['skill\_id']=='2\_37\_48\_77','skill\_id']='2'

data.loc[data['skill\_id']=='2\_48\_79','skill\_id']='2'

data.loc[data['skill\_id']=='2\_70','skill\_id']='2'

data.loc[data['skill\_id']=='21\_22','skill\_id']='21'

data.loc[data['skill\_id']=='24\_46','skill\_id']='24'

data.loc[data['skill\_id']=='27\_321','skill\_id']='27'

data.loc[data['skill\_id']=='292\_293','skill\_id']='292'

data.loc[data['skill\_id']=='34\_69','skill\_id']='34'

data.loc[data['skill\_id']=='34\_81','skill\_id']='34'

data.loc[data['skill\_id']=='34\_278','skill\_id']='34'

data.loc[data['skill\_id']=='35\_46','skill\_id']='35'

data.loc[data['skill\_id']=='46\_79','skill\_id']='46'

data.loc[data['skill\_id']=='47\_63\_70\_77','skill\_id']='47'

data.loc[data['skill\_id']=='48\_63\_70\_79','skill\_id']='48'

data.loc[data['skill\_id']=='48\_63\_79','skill\_id']='48'

data.loc[data['skill\_id']=='49\_50','skill\_id']='49'

data.loc[data['skill\_id']=='5\_102','skill\_id']='5'

data.loc[data['skill\_id']=='5\_375','skill\_id']='5'

data.loc[data['skill\_id']=='51\_53\_75','skill\_id']='51'

data.loc[data['skill\_id']=='58\_85','skill\_id']='58'

data.loc[data['skill\_id']=='63\_75','skill\_id']='63'

data.loc[data['skill\_id']=='74\_85','skill\_id']='74'

data.loc[data['skill\_id']=='74\_92','skill\_id']='74'

data.loc[data['skill\_id']=='9\_12','skill\_id']='9'

data.loc[data['skill\_id']=='9\_13','skill\_id']='9'

data.loc[data['skill\_id']=='9\_14','skill\_id']='9'

data.loc[data['skill\_id']=='9\_15','skill\_id']='9'

data.loc[data['skill\_id']=='92\_323','skill\_id']='92'

data.loc[data['skill\_id']=='97\_99\_105','skill\_id']='97'

data.loc[data['skill\_id']==307,'skill\_id']='307'

#Repair wrong words in skill names

data.loc[data['skill\_name']=='Effect of Changing Dimensions of a Shape Prportionally','skill\_name']='Effect of Changing Dimensions of a Shape Proportionally'

data.loc[data['skill\_name']=='Parts of a Polyomial, Terms, Coefficient, Monomial, Exponent, Variable','skill\_name']='Parts of a Polynomial Terms Coefficient Monomial Exponent Variable'

data.loc[data['skill\_name']=='Order of Operations +,-,/,\* () positive reals','skill\_name']='Order of Operations addition subtraction division multiplication parentheses positive reals'

data.loc[data['skill\_name']=='D.4.8-understanding-concept-of-probabilities','skill\_name']='Understanding concept of probabilities'

data.loc[data['skill\_name']=='Angles - Obtuse, Acute, and Right','skill\_name']='Angles Obtuse Acute and Right'

For Dataset Assistment 2012-13:

data = data[['user\_id','skill\_id','skill','problem\_id','correct']]

# Replace multiply appears

data.loc[data['skill\_id']==163,'skill\_id']=85

data.loc[data['skill\_id']==613,'skill\_id']=590

data.loc[data['skill\_id']==612,'skill\_id']=589

data.loc[data['skill\_id']==571,'skill\_id']=39

data.loc[data['skill\_id']==595,'skill\_id']=39

data.loc[data['skill\_id']==578,'skill\_id']=295

data.loc[data['skill\_id']==601,'skill\_id']=295

data.loc[data['skill\_id']==576,'skill\_id']=296

data.loc[data['skill\_id']==600,'skill\_id']=296

data.loc[data['skill\_id']==570,'skill\_id']=297

data.loc[data['skill\_id']==594,'skill\_id']=297

data.loc[data['skill\_id']==579,'skill\_id']=298

data.loc[data['skill\_id']==602,'skill\_id']=298

data.loc[data['skill\_id']==604,'skill\_id']=581

data.loc[data['skill\_id']==174,'skill\_id']=2

data.loc[data['skill\_id']==572,'skill\_id']=40

data.loc[data['skill\_id']==596,'skill\_id']=40

data.loc[data['skill\_id']==599,'skill\_id']=575

data.loc[data['skill\_id']==597,'skill\_id']=573

data.loc[data['skill\_id']==608,'skill\_id']=585

data.loc[data['skill\_id']==588,'skill\_id']=48

data.loc[data['skill\_id']==611,'skill\_id']=48

data.loc[data['skill\_id']==181,'skill\_id']=86

data.loc[data['skill\_id']==580,'skill\_id']=86

data.loc[data['skill\_id']==603,'skill\_id']=86

data.loc[data['skill\_id']==1641,'skill\_id']=218

data.loc[data['skill\_id']==390,'skill\_id']=106

data.loc[data['skill\_id']==586,'skill\_id']=317

data.loc[data['skill\_id']==609,'skill\_id']=317

data.loc[data['skill\_id']==186,'skill\_id']=4

data.loc[data['skill\_id']==610,'skill\_id']=65

data.loc[data['skill\_id']==587,'skill\_id']=65

data.loc[data['skill\_id']==196,'skill\_id']=12

data.loc[data['skill\_id']==198,'skill\_id']=13

data.loc[data['skill\_id']==606,'skill\_id']=583

data.loc[data['skill\_id']==605,'skill\_id']=582

data.loc[data['skill\_id']==607,'skill\_id']=584

data.loc[data['skill\_id']==202,'skill\_id']=92

data.loc[data['skill\_id']==212,'skill\_id']=79

data.loc[data['skill\_id']==213,'skill\_id']=27

data.loc[data['skill\_id']==216,'skill\_id']=15

data.loc[data['skill\_id']==375,'skill\_id']=222

data.loc[data['skill\_id']==226,'skill\_id']=95

data.loc[data['skill\_id']==598,'skill\_id']=574

data.loc[data['skill\_id']==614,'skill\_id']=591

data.loc[data['skill\_id']==569,'skill\_id']=301

data.loc[data['skill\_id']==593,'skill\_id']=301

data.loc[data['skill\_id']==568,'skill\_id']=307

data.loc[data['skill\_id']==592,'skill\_id']=307

#Repair wrong words in skill names

data.loc[data['skill']=='Effect of Changing Dimensions of a Shape Prportionally','skill']='Effect of Changing Dimensions of a Shape Proportionally'

data.loc[data['skill']=='Parts of a Polyomial, Terms, Coefficient, Monomial, Exponent, Variable','skill']='Parts of a Polynomial Terms Coefficient Monomial Exponent Variable'

data.loc[data['skill']=='Order of Operations +,-,/,\* () positive reals','skill']='Order of Operations addition subtraction division multiplication parentheses positive reals'

data.loc[data['skill']=='D.4.8-understanding-concept-of-probabilities','skill']='Understanding concept of probabilities'

data.loc[data['skill']=='Angles - Obtuse, Acute, and Right','skill']='Angles Obtuse Acute and Right'

data.loc[data['skill']=='Expanded, Standard and Word Notation','skill']='Expanded Standard and Word Notation'

data.loc[data['skill']=='Tree Diagrams, Lists for Counting','skill']='Tree Diagrams Lists for Counting'

data.loc[data['skill']=='Finding y-intercept from Linear Equation','skill']='Finding y intercept from Linear Equation'

data.loc[data['skill']=='Write Linear Equation from Slope and y-intercept','skill']='Write Linear Equation from Slope and y intercept'

data.loc[data['skill']=='X-Y Graph Reading','skill']='X Y Graph Reading'

data.loc[data['skill']=='Comparing and Identifying Slope/Rate of Change','skill']='Comparing and Identifying Slope Rate of Change'

data.loc[data['skill']=='Writine Expression from Diagrams','skill']='Writing Expression from Diagrams'

data.loc[data['skill']=='Calculation with + - \* /','skill']='Calculation with addition subtraction multiplication division'

data.loc[data['skill']=='Line of Best-Fit','skill']='Line of Best Fit'

data.loc[data['skill']=='Properties and Clasification of Pyramid','skill']='Properties and Classification of Pyramid'

data.loc[data['skill']=='Mean-Median-Mode-Range Differentiation','skill']='Mean Median Mode Range Differentiation'

data.loc[data['skill']=='Co-ordinate Points','skill']='Coordinate Points'

data.loc[data['skill']=='Circumference ','skill']='Circumference'

data.loc[data['skill']=='Pattern Finding ','skill']='Pattern Finding'