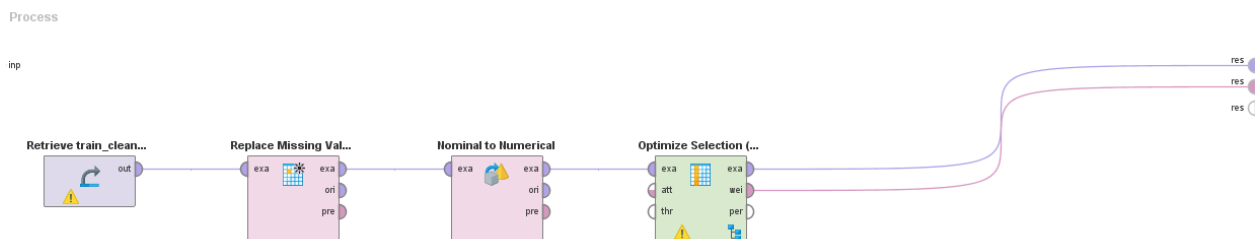


❖ שם הקבוצה ב- Kaggle : **Liad&Michael**

❖ עיבוד מקדים של הנתונים בוצע בתוכנת ה- Rapid Miner (גרסה 7.6) על קובץ האימון והבדיקה וכלל את הפעולות הבאות :

- **השלמת ערכים חסרים** (Missing Values) באמצעות אופרטור Replace Missing Values. ערכים חסרים של תכונות נומריות הושלמו ע"י הממוצע של אותה תכונה, וערכים חסרים של תכונות נומינליות הוחלפו בערך השכיח ביותר.
- **הפיכת תכונות נומינליות לתכונות נומריות** על מנת שניתן יהיה להפעיל אלגוריתמי רגרסיה באמצעות אופרטור Nominal To Numeric (הפיכה ל- Dummy Coding).
- מאחר שקיימים ערכים המופיעים בקובץ האימון ולא בקובץ הבדיקה, הגדרנו את אותם התכונות (Dummy coding) בשני הקבצים.
- הבחנה בין משתנים נומינליים רגילים למשתנים אורדינליים ושמירת ערכי המשתנה על פי הסדר המוגדר ב- Data description.
- **Feature selection** באמצעות האופרטור Optimize Selection (Evolutionary) עם מודל רגרסיה פשוט כדי למזער את מדד ה-RMSE ולהוריד את המימדיות של הנתונים. לאחר ביצוע הפעולה התקבל סט של 100 תכונות המשמעותיות ביותר.
- מצורף צילום המסך של התהליך הראשי (קיימים תתי-תהליכים נוספים) להכנת הנתונים ב- RapidMiner :



המשך בעמוד הבא

מטלה 1 - House Prices

ישום אלגוריתמים לומדים 2017-2018

טבלת התוצאות מ-Kaggle:

איטרציה	השיטה (לכל השיטות בוצע העיבוד המקדים המתואר לעיל)	Score by Kaggle
1	Feature Selection up to 150 attributes + Linear Regression	0.19086
2	Feature Selection up to 150 attributes + ElasticNet Reggression	0.42433
3	Feature Selection up to 150 attributes + GradientBoosting (criterion='friedman_mse', max_depth=3, alpha=0.9)	0.16082
4	Feature Selection up to 150 attributes + GradientBoosting (criterion='friedman_mse', max_depth=5, alpha=0.9)	0.15538
5	Feature Selection up to 150 attributes + GradientBoosting (criterion='friedman_mse', max_depth=8, alpha=0.9)	0.16522
6	Feature Selection up to 100 attributes + GradientBoosting (criterion='friedman_mse', max_depth=5, alpha=0.9)	0.15484
7	Feature Selection up to 100 attributes + GradientBoosting (criterion='friedman_mse', max_depth=5, alpha=0.9) + Warm_Start=True (reuse the solution of the previous calls)	0.15568
8	Feature Selection up to 100 attributes + GradientBoosting (criterion='friedman_mse', max_depth=5, alpha=0.9) + minimise loss function of "Huber"	0.16067
9	Feature Selection up to 100 attributes + Keras Neural Network with 2 hidden layers with 64 and 32 units each (MAE optimization metric and 10 epochs)	1.68804
10	Feature Selection up to 100 attributes + Keras Neural Network with 2 hidden layers with 72 and 16 units each (MAE optimization metric and 50 epochs)	0.46058
11	Feature Selection up to 100 attributes + Keras Neural Network with 4 hidden layers with 100,70,50 and 16 units each (MAE optimization metric and 100 epochs)	0.19364
12	Feature Selection up to 100 attributes + Keras Neural Network with 5 hidden layers with 100,70,50,16 and 5 units each (MAE optimization metric and 100 epochs)	0.20300
13	Feature Selection up to 100 attributes + Ridge regression with built-in cross-validation (alpha=0.1,1.0,10.0)	0.18613

House Prices - 1 מטלה

ישום אלגוריתמים לומדים 2017-2018

0.23641	Feature Selection up to 100 attributes + Lasso model fit with Least Angle Regression (alpha=0.1)	14
0.23669	Feature Selection up to 100 attributes + DecisionTreeRegressor (max_depth=5)	15
0.23987	Feature Selection up to 100 attributes + KNeighborsRegressor (n_neighbour=5)	16
0.18088	Feature Selection up to 100 attributes + Ensemble of: Ensemble - LinearRegression, Ridge, RidgeCV, Lasso (Averaged)	17
0.18590	Feature Selection up to 100 attributes + Ensemble - Lasso, LassoLars, DecisionTreeRegressor, KNeighborsRegressor, ElasticNet (Averaged)	18
0.21894	Feature Selection up to 100 attributes + Ensemble - Ensemble - ElasticNet, GradientBoostingRegressor, NeuralNetwork (Averaged)	19
0.17421	Feature Selection up to 100 attributes + Ensemble - GradientBoostingRegressor, RandomForestRegressor, ExtraTreesRegressor, NeuralNetwork (Average)	20
0.19857	Feature Selection up to 100 attributes + Ensemble - GradientBoostingRegressor, RandomForestRegressor, ExtraTreesRegressor, NeuralNetwork (Min)	21
0.17386	Feature Selection up to 100 attributes + Ensemble - GradientBoostingRegressor, RandomForestRegressor, ExtraTreesRegressor, XGBoostRegressor (Average)	22
0.15055	Feature Selection up to 100 attributes + XGBoostRegressor with GridSearchCV (with n_estimators=200)	23
0.1494	Feature Selection up to 100 attributes + XGBoostRegressor with GridSearchCV (on max_depth= [3, 4, 5] and n_estimators= [50, 100, 200, 500, 1000])	24
0.14845	Feature Selection up to 100 attributes + XGBoostRegressor with GridSearchCV (on max_depth= [3, 4, 5, 6] and learning_rate=0.1-0.6 and n_estimators= [50, 100, 200, 500, 1000])	25
0.14757	Feature Selection up to 100 attributes + XGBoostRegressor with GridSearchCV (on max_depth= [3, 4, 5, 6] and n_estimators= [50, 100, 200, 500, 1000])	26