Approach Machinehack analytics olympiad 2022

Create a machine learning model to help an insurance company understand which claims are worth rejecting and the claims which should be accepted for reimbursement.

- Basic exploratory data analysis using pandas, matplotlib, seaborn packages.
- Data pre-processing
    - Column name lower casr
    - Feature Engineering
        - Create a binary indicator to indicate if the policyholder has a child without marriage.
        - get a categorical columns percentage
        - numerical groupby numerical summary
- The final features for the model

    0_age

    1_gender

    2_driving_experience

    3_education

    4_income

    5_credit_score

    6_vehicle_ownership

    7_vehicle_year

    8_married

    9_children

10_annual_mileage

11_speeding_violations

12_duis

13_past_accidents

14_type_of_vehicle

15_not_married_hv_child

16_id_age_percentage

17_id_gender_percentage

18_id_driving_experience_percentage

19_id_education_percentage

20_id_income_percentage

21_id_vehicle_ownership_percentage

22_id_vehicle_year_percentage

23_id_married_percentage

24_id_children_percentage

25_id_type_of_vehicle_percentage

26_id_not_married_hv_child_percentage

27_id_annual_mileage_percentage

28_id_speeding_violations_percentage

29_id_duis_percentage

30_id_past_accidents_percentage

31_age_credit_score_median

32_age_credit_score_min

33_age_credit_score_mean

34_age_credit_score_max

35_age_annual_mileage_median

36_age_annual_mileage_min

37_age_annual_mileage_mean

38_age_annual_mileage_max

39_age_speeding_violations_median

40_age_speeding_violations_min

41_age_speeding_violations_mean

42_age_speeding_violations_max

43_age_duis_median

44_age_duis_min

45_age_duis_mean

46_age_duis_max

47_age_past_accidents_median

48_age_past_accidents_min

49_age_past_accidents_mean

50_age_past_accidents_max

51_gender_credit_score_median

52_gender_credit_score_min

53_gender_credit_score_mean

54_gender_credit_score_max

55_gender_annual_mileage_median

56_gender_annual_mileage_min

57_gender_annual_mileage_mean

58_gender_annual_mileage_max

59_gender_speeding_violations_median

60_gender_speeding_violations_min

61_gender_speeding_violations_mean

62_gender_speeding_violations_max

63_gender_duis_median

64_gender_duis_min

65_gender_duis_mean

66_gender_duis_max

67_gender_past_accidents_median

68_gender_past_accidents_min

69_gender_past_accidents_mean

70_gender_past_accidents_max

71_driving_experience_credit_score_median

72_driving_experience_credit_score_min

73_driving_experience_credit_score_mean

74_driving_experience_credit_score_max

75_driving_experience_annual_mileage_median

76_driving_experience_annual_mileage_min

77_driving_experience_annual_mileage_mean

78_driving_experience_annual_mileage_max

79_driving_experience_speeding_violations_median

80_driving_experience_speeding_violations_min

81_driving_experience_speeding_violations_mean

82_driving_experience_speeding_violations_max

83_driving_experience_duis_median

84_driving_experience_duis_min

85_driving_experience_duis_mean
86_driving_experience_duis_max
87_driving_experience_past_accidents_median
88_driving_experience_past_accidents_min
89_driving_experience_past_accidents_mean
90_driving_experience_past_accidents_max
91_education_credit_score_median
92_education_credit_score_min
93_education_credit_score_mean
94_education_credit_score_max
95_education_annual_mileage_median
96_education_annual_mileage_min
97_education_annual_mileage_mean
98_education_annual_mileage_max
99_education_speeding_violations_median
100_education_speeding_violations_min
101_education_speeding_violations_mean
102_education_speeding_violations_max
103_education_duis_median
104_education_duis_min
105_education_duis_mean
106_education_duis_max
107_education_past_accidents_median
108_education_past_accidents_min
109_education_past_accidents_mean

110_education_past_accidents_max

111_income_credit_score_median

112_income_credit_score_min

113_income_credit_score_mean

114_income_credit_score_max

115_income_annual_mileage_median

116_income_annual_mileage_min

117_income_annual_mileage_mean

118_income_annual_mileage_max

119_income_speeding_violations_median

120_income_speeding_violations_min

121_income_speeding_violations_mean

122_income_speeding_violations_max

123_income_duis_median

124_income_duis_min

125_income_duis_mean

126_income_duis_max

127_income_past_accidents_median

128_income_past_accidents_min

129_income_past_accidents_mean

130_income_past_accidents_max

131_vehicle_ownership_credit_score_median

132_vehicle_ownership_credit_score_min

133_vehicle_ownership_credit_score_mean

134_vehicle_ownership_credit_score_max

135_vehicle_ownership_annual_mileage_median

136_vehicle_ownership_annual_mileage_min

137_vehicle_ownership_annual_mileage_mean

138_vehicle_ownership_annual_mileage_max

139_vehicle_ownership_speeding_violations_median

140_vehicle_ownership_speeding_violations_min

141_vehicle_ownership_speeding_violations_mean

142_vehicle_ownership_speeding_violations_max

143_vehicle_ownership_duis_median

144_vehicle_ownership_duis_min

145_vehicle_ownership_duis_mean

146_vehicle_ownership_duis_max

147_vehicle_ownership_past_accidents_median

148_vehicle_ownership_past_accidents_min

149_vehicle_ownership_past_accidents_mean

150_vehicle_ownership_past_accidents_max

151_vehicle_year_credit_score_median

152_vehicle_year_credit_score_min

153_vehicle_year_credit_score_mean

154_vehicle_year_credit_score_max

155_vehicle_year_annual_mileage_median

156_vehicle_year_annual_mileage_min

157_vehicle_year_annual_mileage_mean

158_vehicle_year_annual_mileage_max

159_vehicle_year_speeding_violations_median

160_vehicle_year_speeding_violations_min

161_vehicle_year_speeding_violations_mean

162_vehicle_year_speeding_violations_max

163_vehicle_year_duis_median

164_vehicle_year_duis_min

165_vehicle_year_duis_mean

166_vehicle_year_duis_max

167_vehicle_year_past_accidents_median

168_vehicle_year_past_accidents_min

169_vehicle_year_past_accidents_mean

170_vehicle_year_past_accidents_max

171_married_credit_score_median

172_married_credit_score_min

173_married_credit_score_mean

174_married_credit_score_max

175_married_annual_mileage_median

176_married_annual_mileage_min

177_married_annual_mileage_mean

178_married_annual_mileage_max

179_married_speeding_violations_median

180_married_speeding_violations_min

181_married_speeding_violations_mean

182_married_speeding_violations_max

183_married_duis_median

184_married_duis_min

185_married_duis_mean

186_married_duis_max

187_married_past_accidents_median

188_married_past_accidents_min

189_married_past_accidents_mean

190_married_past_accidents_max

191_not_married_hv_child_credit_score_median

192_not_married_hv_child_credit_score_min

193_not_married_hv_child_credit_score_mean

194_not_married_hv_child_credit_score_max

195_not_married_hv_child_annual_mileage_median

196_not_married_hv_child_annual_mileage_min

197_not_married_hv_child_annual_mileage_mean

198_not_married_hv_child_annual_mileage_max

199_not_married_hv_child_speeding_violations_median

200_not_married_hv_child_speeding_violations_min

201_not_married_hv_child_speeding_violations_mean

202_not_married_hv_child_speeding_violations_max

203_not_married_hv_child_duis_median

204_not_married_hv_child_duis_min

205_not_married_hv_child_duis_mean

206_not_married_hv_child_duis_max

207_not_married_hv_child_past_accidents_median

208_not_married_hv_child_past_accidents_min

209_not_married_hv_child_past_accidents_mean

210_not_married_hv_child_past_accidents_max

211_children_credit_score_median

212_children_credit_score_min

213_children_credit_score_mean

214_children_credit_score_max

215_children_annual_mileage_median

216_children_annual_mileage_min

217_children_annual_mileage_mean

218_children_annual_mileage_max

219_children_speeding_violations_median

220_children_speeding_violations_min

221_children_speeding_violations_mean

222_children_speeding_violations_max

223_children_duis_median

224_children_duis_min

225_children_duis_mean

226_children_duis_max

227_children_past_accidents_median

228_children_past_accidents_min

229_children_past_accidents_mean

230_children_past_accidents_max

231_type_of_vehicle_credit_score_median

232_type_of_vehicle_credit_score_min

233_type_of_vehicle_credit_score_mean

234_type_of_vehicle_credit_score_max

   - 235_type_of_vehicle_annual_mileage_median
   - 236_type_of_vehicle_annual_mileage_min
   - 237_type_of_vehicle_annual_mileage_mean
   - 238_type_of_vehicle_annual_mileage_max
   - 239_type_of_vehicle_speeding_violations_median
   - 240_type_of_vehicle_speeding_violations_min
   - 241_type_of_vehicle_speeding_violations_mean
   - 242_type_of_vehicle_speeding_violations_max
   - 243_type_of_vehicle_duis_median
   - 244_type_of_vehicle_duis_min
   - 245_type_of_vehicle_duis_mean
   - 246_type_of_vehicle_duis_max
   - 247_type_of_vehicle_past_accidents_median
   - 248_type_of_vehicle_past_accidents_min
   - 249_type_of_vehicle_past_accidents_mean
   - 250_type_of_vehicle_past_accidents_max
- Created catboost classifier model and tuned hyperparameters by using optuna framework.Model evaluated with Logloss.After 100 trials,
   - The best score is 0. 680893

- The best hyperparemeters are,

  {'reg_lambda': 605,

  'learning_rate': 0.04731571585972637,

  'n_estimators': 772,

  'max_depth': 6,

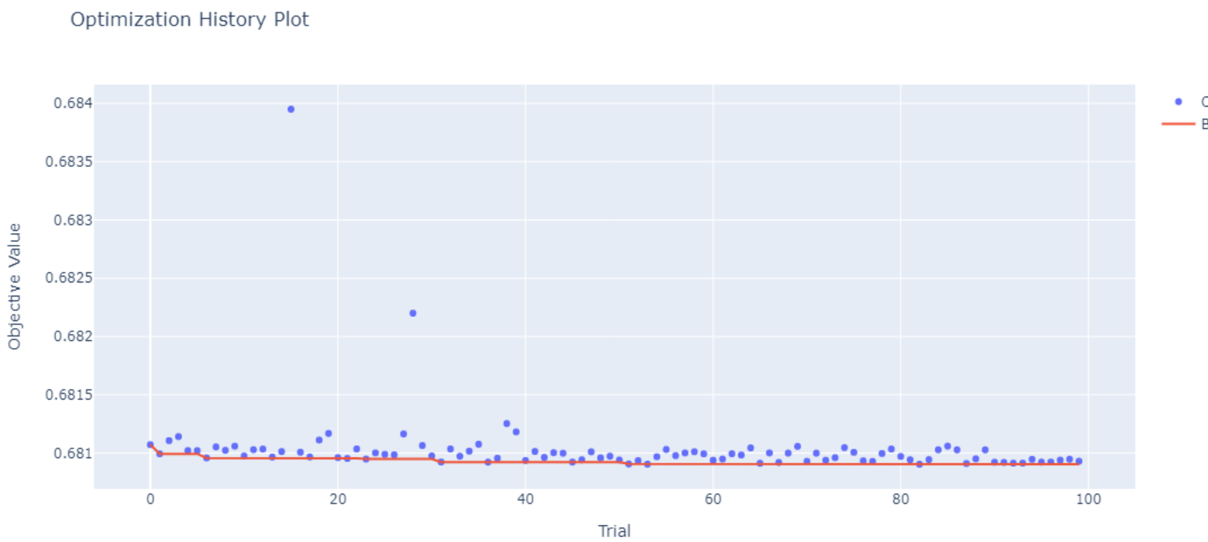
  'random_state': 1024,

  'boosting_type': 'Plain',

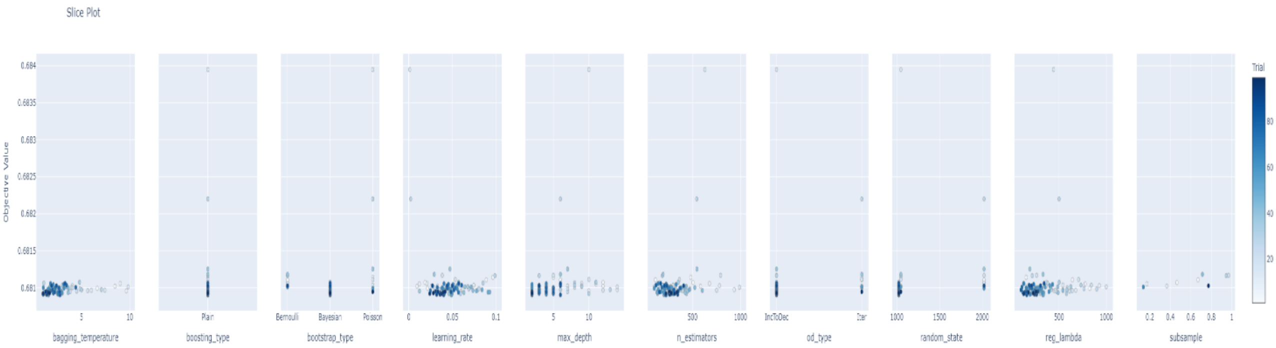
  'bootstrap_type': 'Bayesian',

  'bagging_temperature': 8.801547913767672}

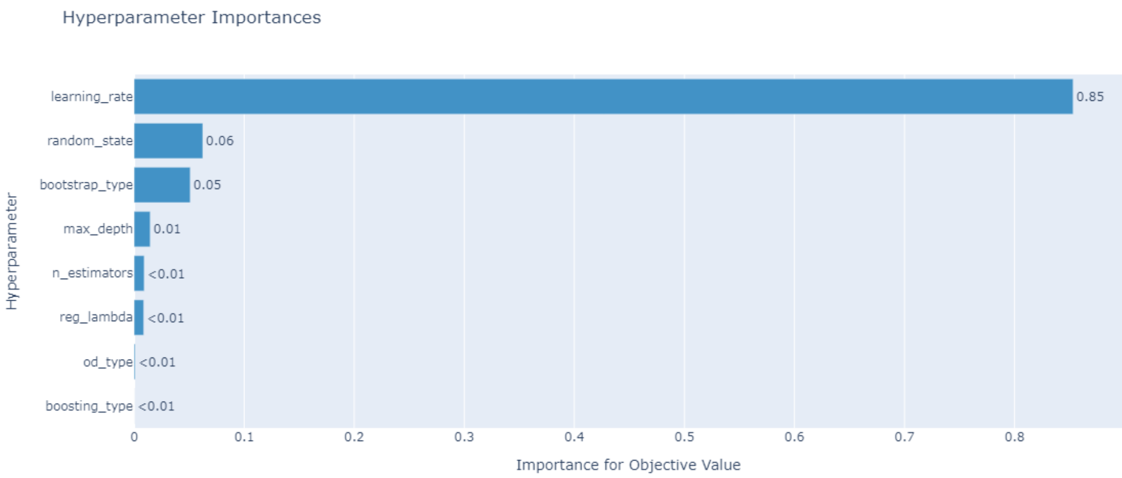
- Visualizing the Optimization History - Explains the best score at each trials.



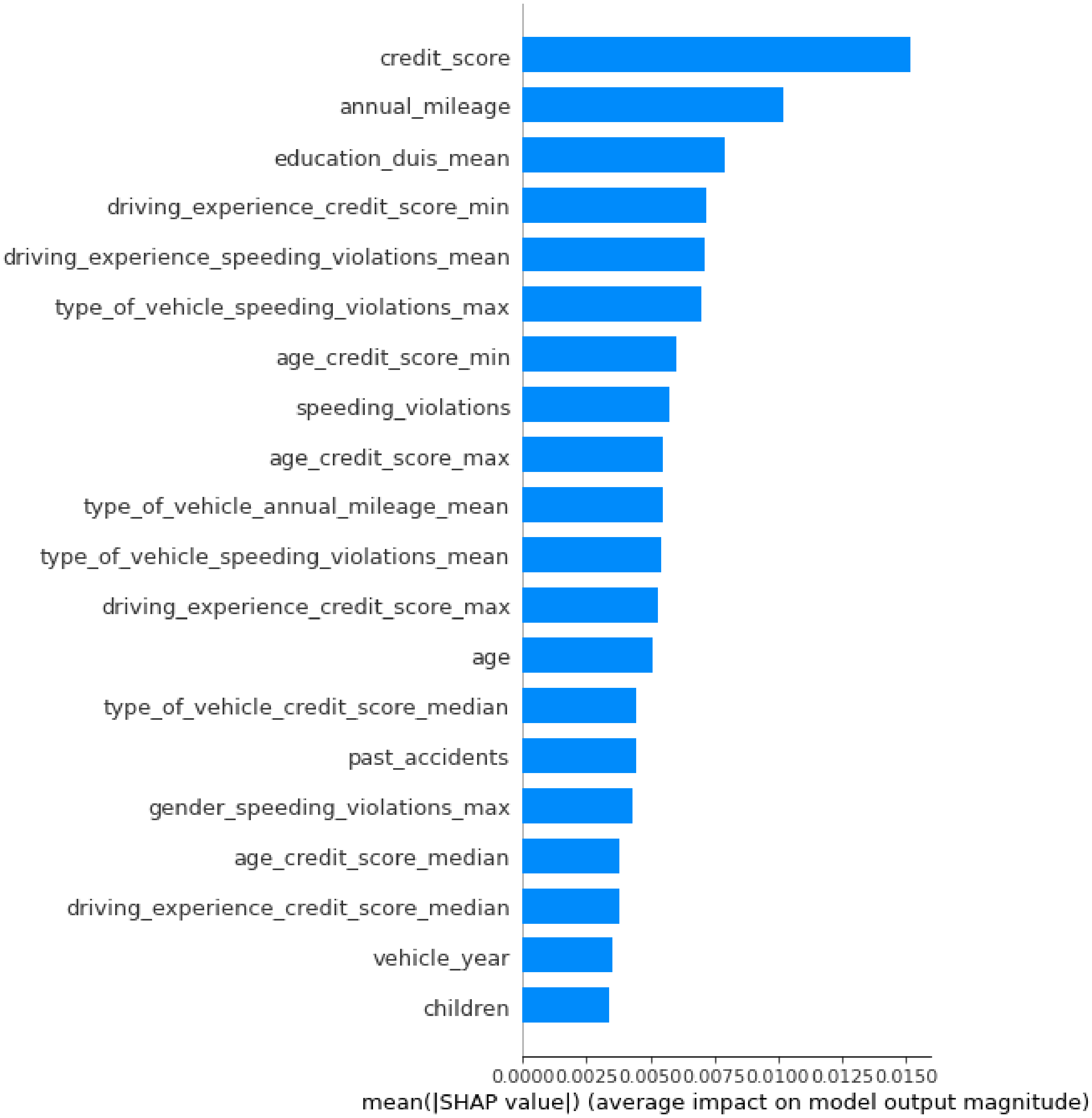
- Visualizing High-dimensional Parameter Relationships



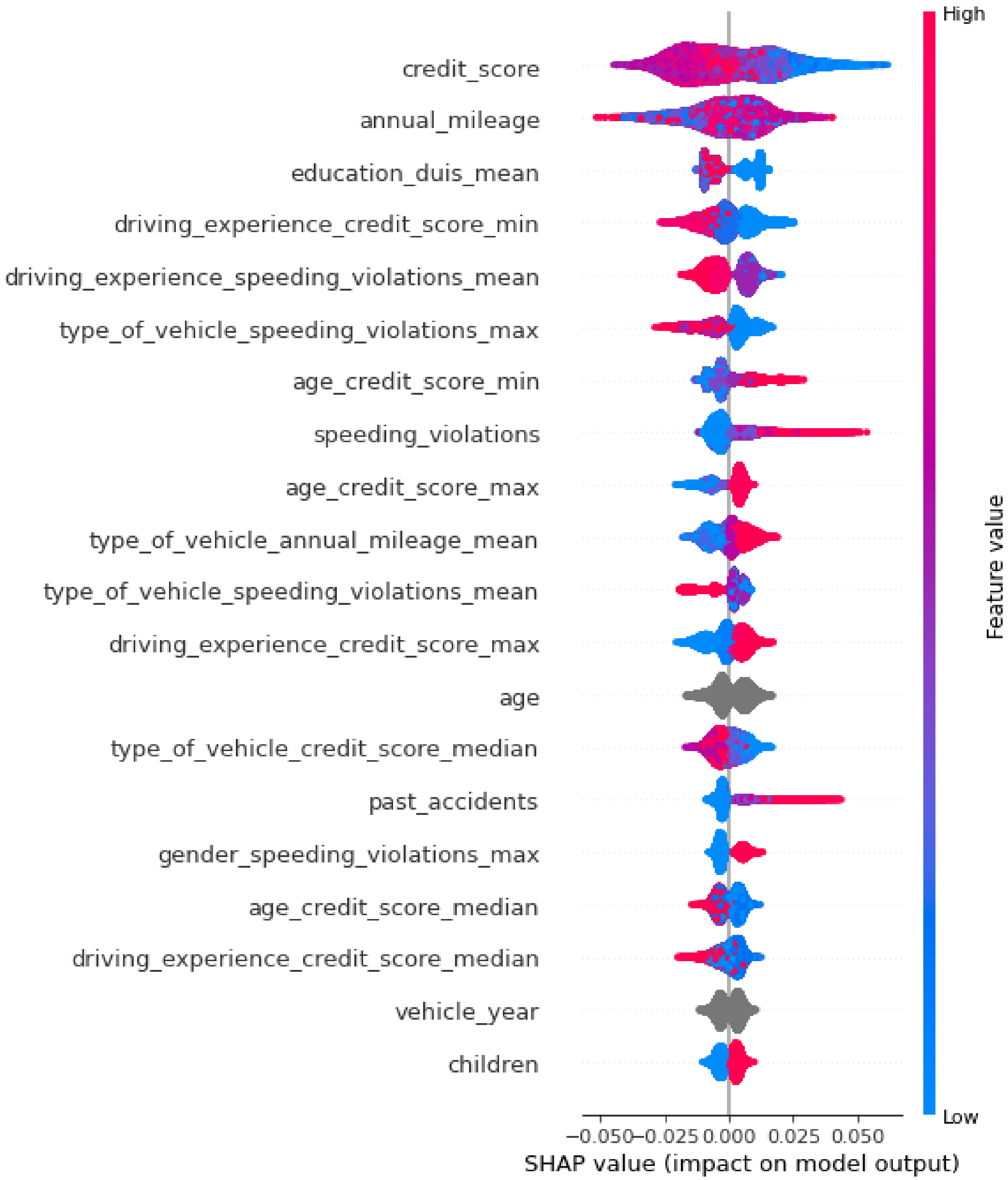
- Visualizing Parameter Importances



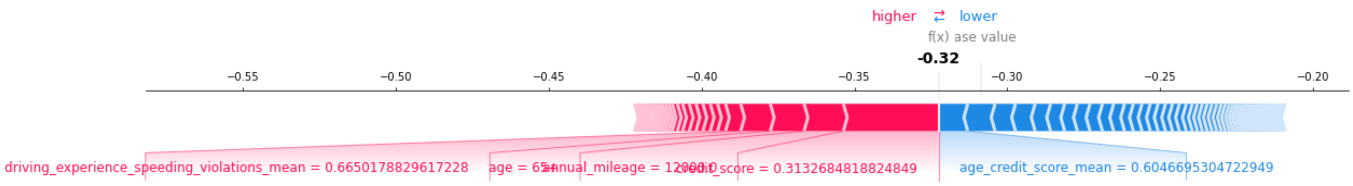
- Catboost SHAP feature importances plot



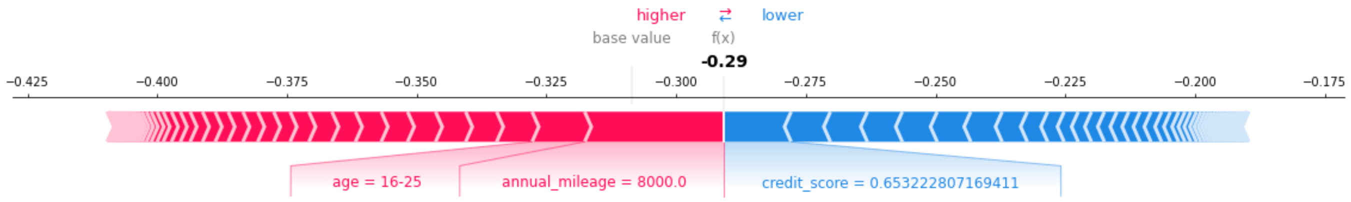
- Catboost SHAP top features impact the model



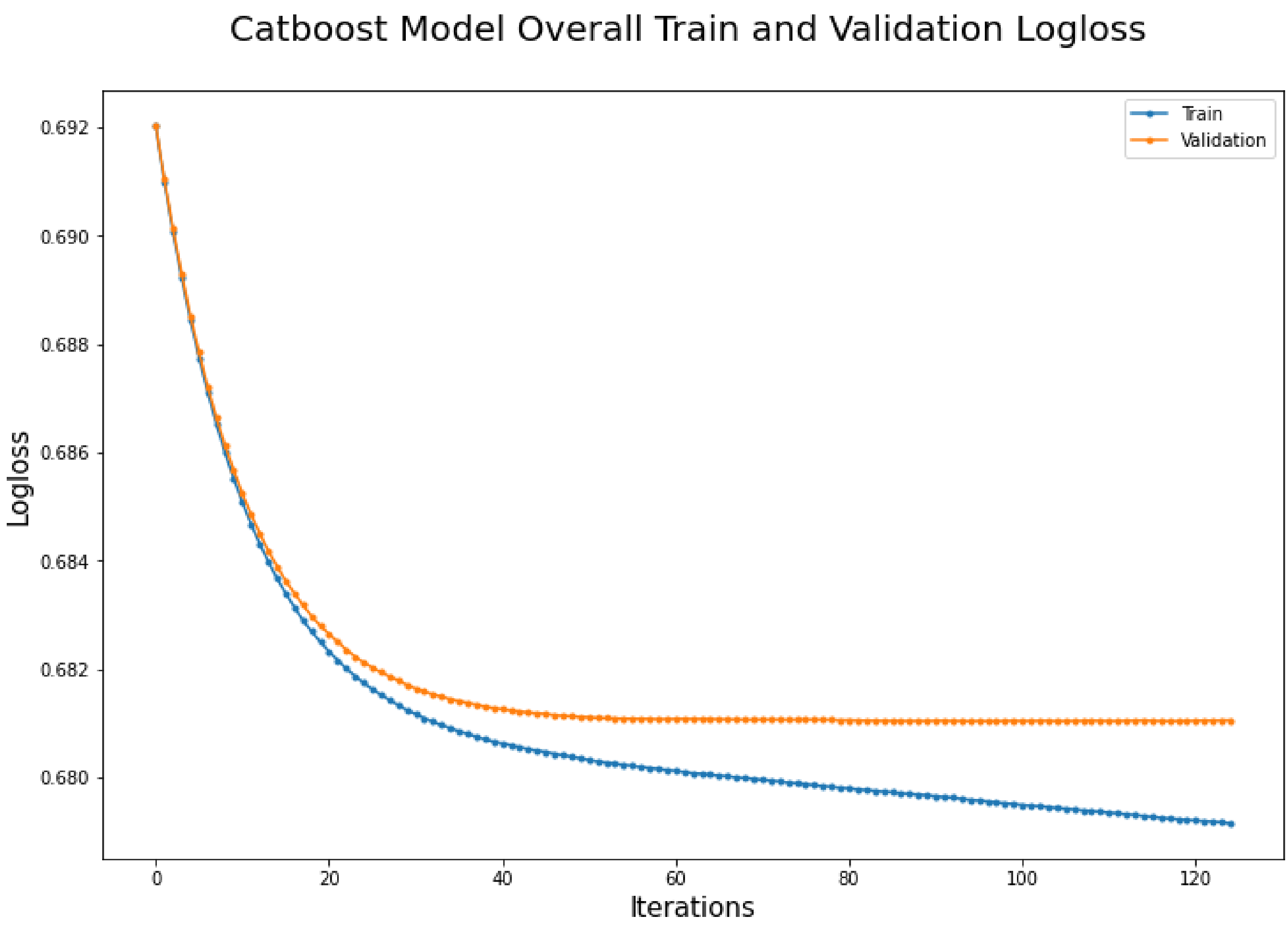
- Top feature influences for class 1



- Top feature influences for class 0



- Overall Train and Validation Logloss



- Train logloss: 0.67978 , Validation logloss: 0.68106
- Final competition score
    - Private LB: 0.68081
    - Public LB: 0.68037