Aims
To identify the predictors of adolescents smoking behavior in South Korea using a machine-learning approach

Introduction
• Smoking is known as a modifiable risk behavior that causes various health problems including cancer
• Globally, adolescent smoking is likely to persist through adulthood
• In South Korea, smoking behaviors among Korean adolescents remains a significant social problem
• For an effective intervention, the factors that underlie and influence the behavior of smoking should be identified.
• Machine learning is an approach that is well suited to reveal patterns of information in large, complex datasets that are useful in predicting outcomes (Chekroud, 2016)

Method
• Data
  ➢ The 2015 Korean Youth Risk Behaviors Web-based Survey (KYRBS) was used.
  ➢ Data completed items concerning smoking were used.
  ➢ Therefore, data from 5,123 among 68,043 students in grades 7 through 12 were used.
• Machine-learning approach
  ➢ Machine-learning pipeline developed by Fayyad (1996) and Yoon (2015) was applied.
  ➢ Clinically meaningful features based on the conceptual framework for adolescent risk behaviors (Jessor, 1991) were selected for analysis.
  ➢ This process was expected to reduce the “curse of dimensionality” (a high number of inter-related variables in large dataset interfere with the accuracy of the machine-learning model)
  ➢ Three machine learning algorithms embedded in Weka (i.e., J48, Naïve Bayes, and Logistic Regression) were applied.

Framework

<table>
<thead>
<tr>
<th>Biology / Genetics</th>
<th>Social environment</th>
<th>Perceived environment</th>
<th>Personality</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Family history of smoking</td>
<td>• SES (socioeconomic status)</td>
<td>• Best friends or teachers' smoking behavior</td>
<td>• Depression</td>
<td>• Alcohol consumption</td>
</tr>
<tr>
<td>• School achievement</td>
<td>• Exposure to anti-smoking advertising campaigns</td>
<td>• Smoking cessation education in school</td>
<td>• Suicidal ideation</td>
<td>• Physical activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Perceived health status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Perceived stress level</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results
• Majority of the adolescents in this study were male and high school students (80% and 78%, respectively)
• The Logistic Regression algorithm demonstrated the highest level of accuracy (84.0%, F-measure = 0.795).
• Grade (-0.06) and alcohol consumption (-0.56) were the top two features with the highest coefficients.
  ➢ Middle school students and students who had never drank alcohol were highly associated with the behavior of smoking

Conclusions
• This study demonstrates the behavioral predictors associated with smoking using the KYRBY.
• The results were inconsistent with the previous studies.
• Further study with association between smoking behaviors and alcohol consumption among Korean adolescent is needed

References

Acknowledgments
Funded by the University Of Ulsan.