Stream Analysis of Incoming Events Using Different Data Analysis Methods

Bc. Matúš Cimerman, Dr. Jakub Ševcech

STU
FIIT

SLOVAK UNIVERSITY OF
TECHNOLOGY IN BRATISLAVA
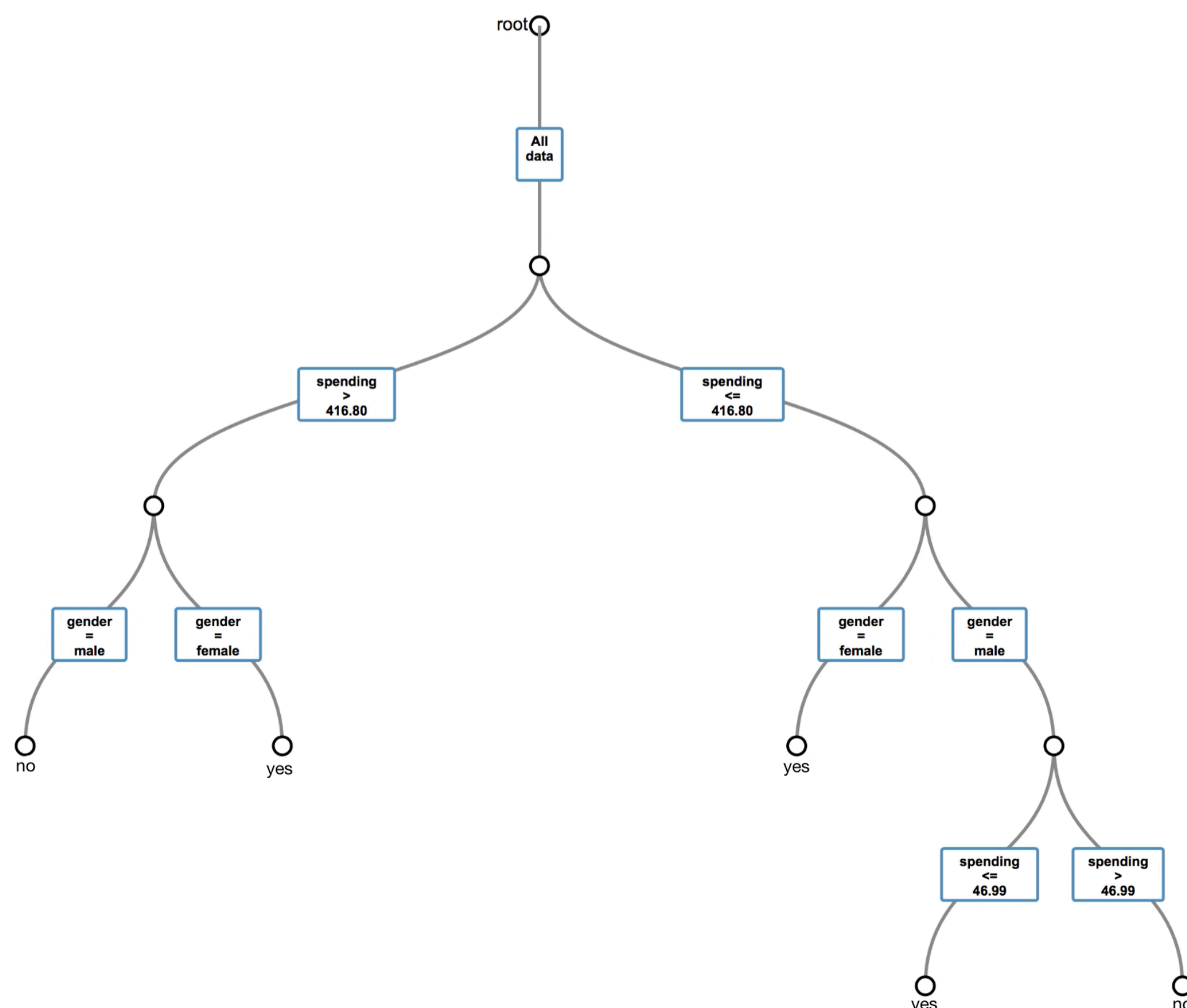FACULTY OF INFORMATICS
AND INFORMATION TECHNOLOGIES

Goals:

• Real-time animated model visualization.
• Ease of model interpretation for data domain experts without machine learning algorithms knowledge.
• Model is adapting to drifts and visualizing drift adaptation.
• Accuracy of used classification method is comparable to similar approaches.
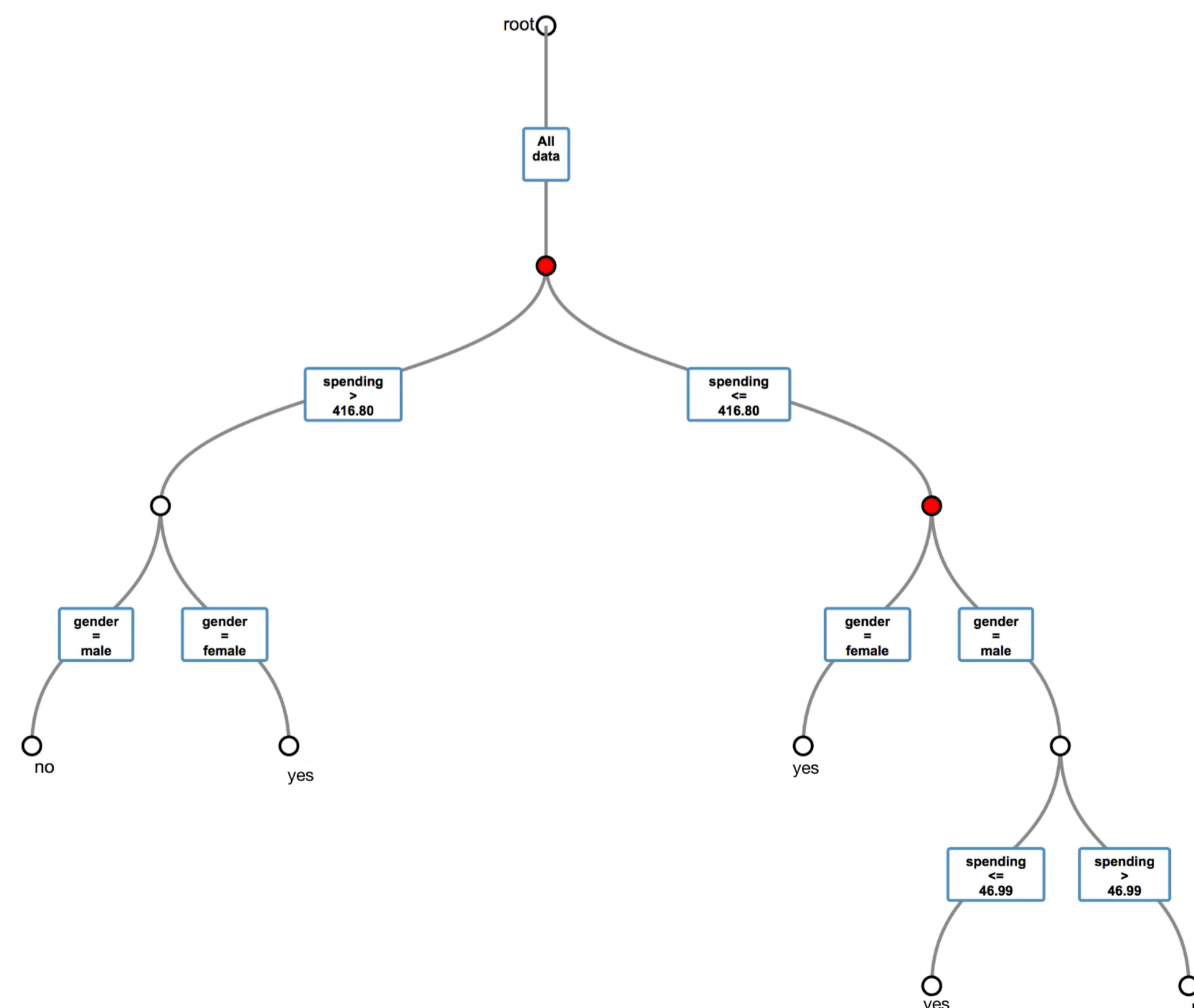
Hoeffding adaptive trees:

• We are using Hoeffding trees with ADWIN adaptive sliding window in task of classification.
• Method successfully discover drifts and adapts to them.
• Tree consists of learning and split nodes. Learning nodes are leafs which can split. Split nodes are also drift detectors.
• When drift is detected, alternating tree starts growing.
• If alternating tree has lower error rate, it replaces its origin or is erased.
• Hoeffding bound is used to measure split and alternating trees confidence.
• Achieved accuracy using artificial dataset was 0.89 and Air dataset was 0.656.
• In qualitative experiment we evaluated ability to identify drifts, all participants precisely identified drift influencing tree model.

Discovering drift during real-time training:

1. Tree in stable phase   2. Drift was discovered   3. Old tree is being replaced with alternating tree