Assignment 4

Temesgen Mehari, Oliver Philipps, Stefan Vikoler

Data Engineering

16. December 2015

Task I: Business understanding

- Business objectives
  - What is the problem?
    - Relate performance to software configuration.
    - Better quality and efficient use of resources.
    - Cost for database.
  - Who has the problem?
    - Product managers and software architects.
  - Problem measurement
    - Increasing database performance in milliseconds.
- Business success criteria
  - Better performance and efficient use of resources.
  - Reduce hardware cost for database.
  - Faster systems.
  - Increase competition on the market and get new customers.
  - Making decisions.

Task I: Business understanding

- Determine data mining goals
  - Find the right configuration for performance increasement.
- Data mining success criteria
  - Measure realtime preformance with predicted performance.

Task II: Data understanding

- 2560 instances with 19 fields
  - 18 boolean values {Y,N}
  - 1 numeric value
- performance
  - Min: 0,3536114359
  - Max: 39,0850482432
  - Mean: 8,2726768095

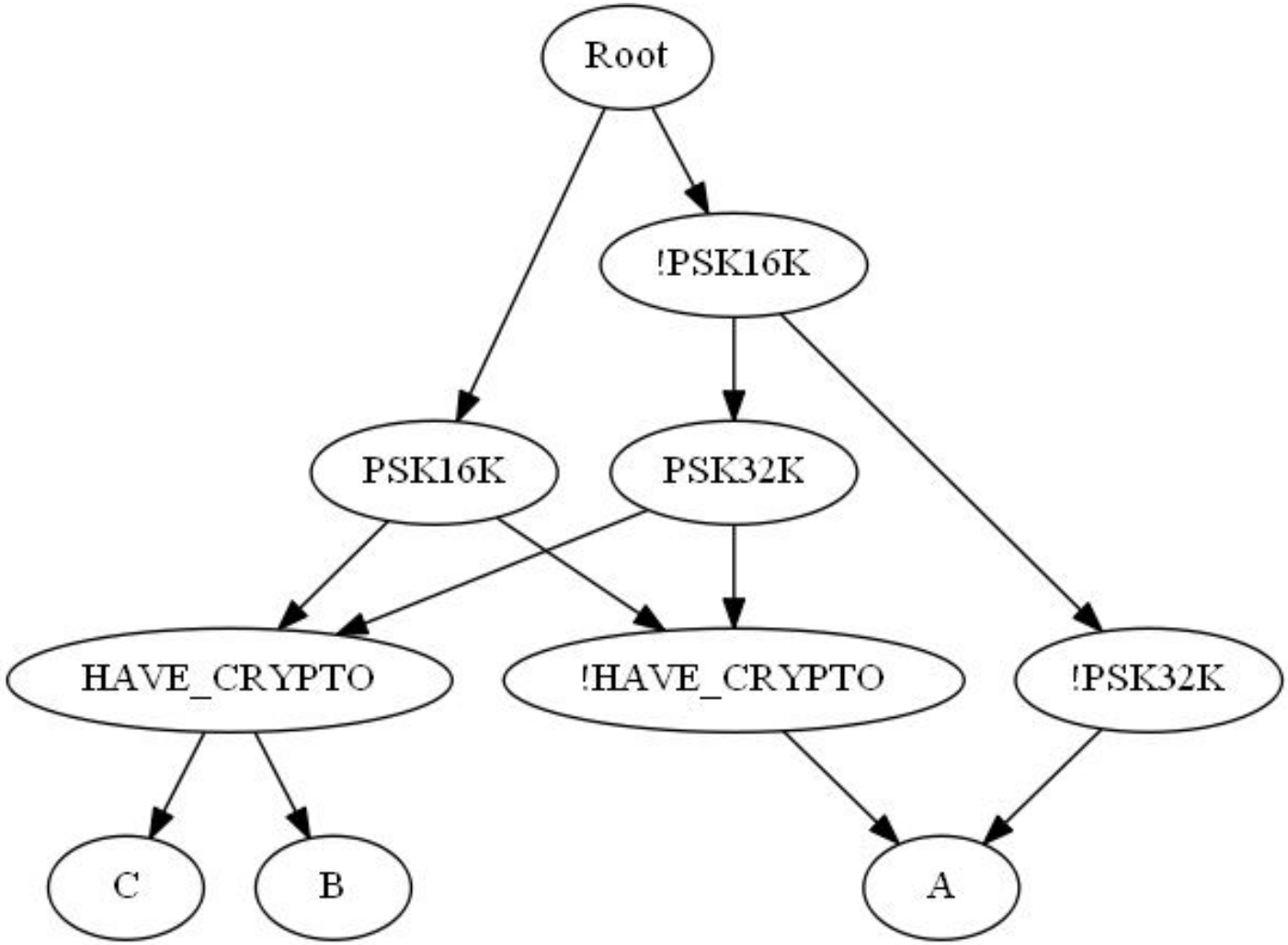
Task III: Data preparation

- Not useful attributes
  - CACHESIZE
  - PAGESIZE
- Performance in groups
  - Group A: performance $<$ 14
  - Group B: 14 $<=$ performance $<=$ 28
  - Group C: performance $>$ 28
- changes of values {Y,N} to {t,f}

Task IV: Modeling

- WEKA
  - BFTree
  - 5 leaves
  - treesize 9

Task V: Evaluation



Task VI: Plan deployment

- Analysis of software architecture
- Measure
  - predicted performance VS real-time performance
  - compare with old products/projects
  - compare hardware costs with old products/projects
- Propagate the knowledge
  - Kick off meeting
  - Meeting with software architects
- Pitfalls
  - You might need some configurations which decrease performance.