Detecting Deceptive Opinion
Spam in Online Reviews

Ivo de Vries – 1820672
Supervisor: Davide Ceolin
Context & Problem Statement

• iens.nl
  – Largest restaurant site in Holland
  – View restaurant information & post reviews
  – Receive 2000 – 3000 reviews each week
  – 2 full time and 3 part time employees manually filter reviews

• Deceptive opinion spam is inherently difficult to detect

• Fraction of deceptive opinion spam is filtered
Research Questions

1. Can we use algorithms to detect deceptive opinion spam in online reviews?
   a) Can meta-data analysis on reviews/reviewers indicate opinion spam?
   b) Can natural language processing on the review content indicate opinion spam?
   c) Can we combine meta-data analysis and natural language processing techniques to increase the accuracy of the spam detection algorithms?
Approach and Method

• 3 entities that are intimately related:
  – Review, Reviewer and Restaurant

• Score each entity on trustworthiness:
  – Restaurant: remove all first time reviewers
  – Reviewer: deviation ratings deviate from average
  – Review: review resemblance

• Data provenance

• Combine scores to evaluate review on spam
Expected Outcome & Contribution

• Scoring system that combines different techniques for detecting review spam
• Automatic spam detection to be at least on par with human evaluators
• Provide the basis for an automated spam detection system