

Whisky-Empfehlungen

Agenda

- Motivation
- Einführung in die Domäne Whisky
- Einführung Empfehlungssysteme
 - Einführung KDD, Data Mining, Clustering
- Aktueller Stand & Aufgaben

Motivation

→ Problem Whisky-Empfehlung

→ große Auswahl

→ objektive Empfehlung kaum möglich

→ Empfehlungssysteme prägen Alltag

Einführung in die Domäne Whisky

→ **Sehr vielfältige Welt**

→ **Aromatisch/Geschmacklich sehr verschiedene Ausprägungen**

→ Von Destillieren so gewünscht – abheben von Konkurrenz

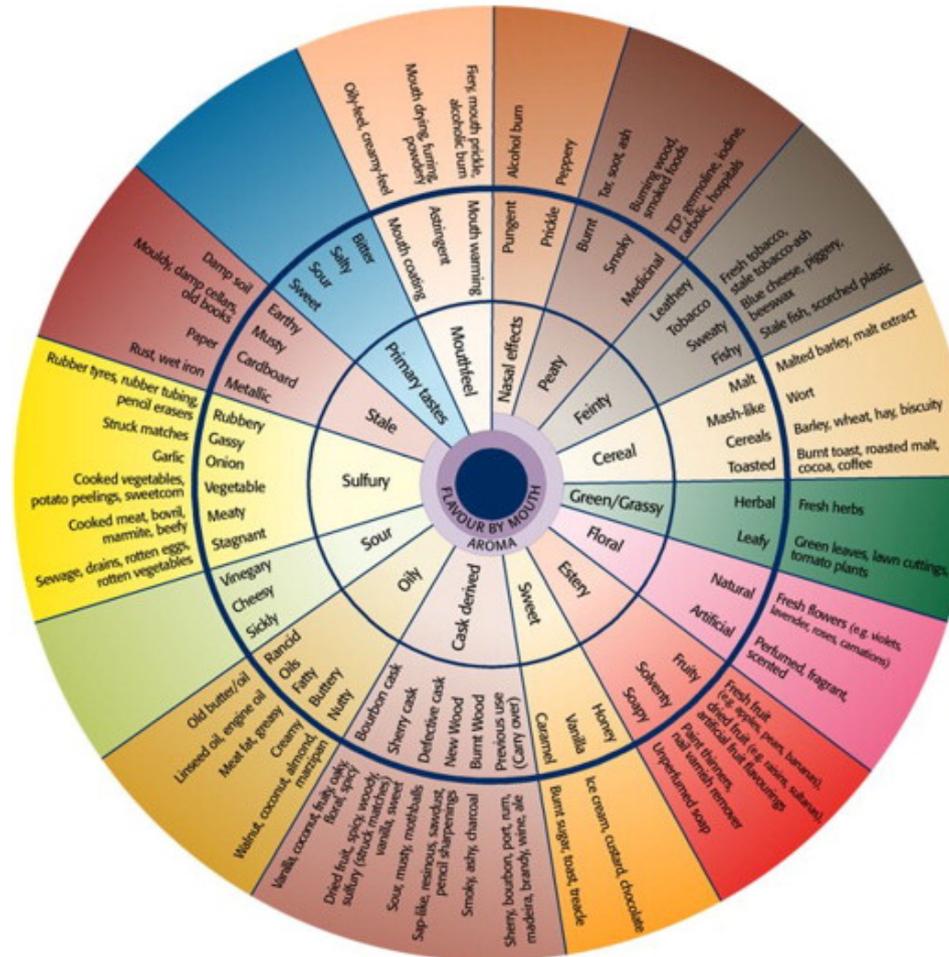
→ **komplexer, langer Herstellungsprozess**

→ keine zusätzlichen Aromastoffe erlaubt

Herstellungsprozess

Produktionsschritt	Auswirkung
Rohmaterial (getorfte Whiskymalz)	Rauch, Malz, Biskuit, ...
Gärung	Ester, Aldehyde, Säuren
Destillation	Manche Aromen werden herausgefiltert, andere wie Getreide und Gras entstehen
Reifung	Fass gibt Farbe und Aromastoffe wie Vanillin und Eichenlaktone (Kokos) ab, einige Aromen verfliegen („Angel's Share“)

Nosing-Wheel



Nosing-Wheel nach (Jack, 2014, S. 238)

Einführung in Empfehlungssysteme

→ **Software-Tools zur Ermittlung interessanter *Items***

→ **Für Anbieter wie Nutzer hilfreich**

→ Anbieter hat in der Regel finanzielle Interessen

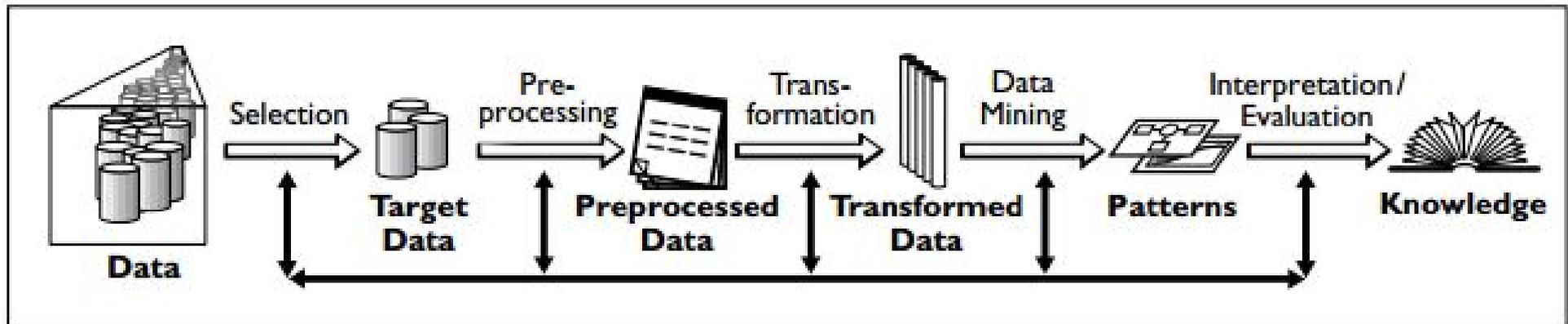
→ Nutzer will schnell das richtige/beste finden

→ **Verschiedene Ansätze**

→ hier vor allem Inhaltsorientierter Ansatz interessant: Items anhand ihrer Ähnlichkeit zueinander empfehlen

→ weitere Ansätze: collaborative filtering, demographisch, wissensbasiert, community based, hybrid

Knowledge Discovery in Databases



Der KDD-Prozess (Fayyad u. a., 1996b, S. 29)

→ Allgemeiner Prozess zur Wissensgewinnung aus Rohdaten

→ Verständnis der Domäne → Auswahl der Rohdaten → Vorverarbeitung der Rohdaten → Transformation in eine Repräsentationsform → Auswahl der Data-Mining-Methoden → Durchführung des Data Mining → Interpretation der ermittelten Muster

→ **Zentraler Bestandteil des KDD-Prozesses**

→ **Wieder verschiedene Methoden**

→ Hier Clustering von Interesse: Ermitteln von Klassen in Rohdaten anhand der Nähe einzelner Datensätze zueinander

→ **Drei wesentliche Bestandteile**

→ Musterrepräsentation beschreibt mögliche zu entdeckende Muster

→ Musterevaluation nennt Kriterien zur Bewertung entdeckter Muster

→ Suchalgorithmus findet geeignetste Muster

Clustering

→ **Data-Mining-Methode zur Klassenbildung**

→ **In mehrere Schritte unterteilt**

→ Festlegung der Datenrepräsentation

→ Definition der Datennähe/-Distanz

→ im einfachsten Fall euklidische Distanz

→ Durchführung des Clustering

→ hart vs. fuzzy

→ hierarchisch vs. partitionell

→ (evtl.) Datenabstraktion zur besseren Lesbarkeit des Ergebnisses

→ Ergebnisbewertung

Aktueller Stand

→ Einarbeitung in entsprechende Technologien

→ Empfehlungssysteme, KDD, Data Mining, Clustering

→ Einarbeitung in Whisky

→ Herstellung, Zusammensetzung

Aktuelle Aufgaben

→ Sichtung von möglichen Datenquellen

- Bücher und Online-Datenbanken mit Tasting-Notes
 - Qualität der Quellen untersuchen
- Eine oder mehrere Quellen?
 - Vereinbarkeit verschiedener Quellen prüfen

→ Auswahl von Methoden zur Extraktion und Transformation

- Bücher auslesen
- Websites auslesen

Mögliche Datenquellen (Auswahl)

→ Bücher

→ Murray, Jim: *Jim Murray's Whisky Bible 2017*. Dram Good Books, 2017

→ Ronde, Ingvar: *Malt Whisky Yearbook 2017*. MapDig Media Limited, 2017

→ Online

→ www.scotchwhisky.com

→ www.whiskymag.com

→ www.whisky-monitor.com

Quellen

[Fayyad u.a. 1996a]

Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic: *From data mining to knowledge discovery in databases*. In: *AI magazine* 17 (1996), Nr. 3, S. 37. –
<http://dx.doi.org/10.1609/aimag.v17i3.1230>

[Fayyad u.a. 1996b]

Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic: *The KDD Process for Extracting Useful Knowledge from Volumes of Data*. In: *Commun. ACM* 39 (1996), November, Nr. 11, S. 27–34. - <http://doi.acm.org/10.1145/240455.240464> – ISSN 0001-0782

[Jack 2014]

Jack, Frances: *Sensory analysis*. In: Russell, Inge (Hrsg.); Stewart, Graham (Hrsg.): *Whisky. Second edition*. San Diego : Academic Press, 2014, S. 229 – 242. –
<http://www.sciencedirect.com/science/article/pii/B9780124017351000131> – ISBN 978-0-12-401735-1

Quellen

[Jain u.a. 1999]

Jain, A. K.; Murty, M. N.; Flynn, P. J.: *Data Clustering: A Review*. In: *ACM Comput. Surv.* 31 (1999), September, Nr. 3, S. 264–323. – <http://doi.acm.org/10.1145/331499.331504>. – ISSN 0360-0300

[Ricci u.a. 2011]

Ricci, Francesco; Rokach, Lior; Shapira, Bracha: *Introduction to Recommender Systems Handbook*. S. 1–35. In: Ricci, Francesco (Hrsg.) ; Rokach, Lior (Hrsg.); Shapira, Bracha (Hrsg.); Kantor, B. P. (Hrsg.): *Recommender Systems Handbook*. Boston, MA : Springer US, 2011. – http://dx.doi.org/10.1007/978-0-387-85820-3_1. – ISBN 978-0-387-85820-3

[Sharafi 2013]

Sharafi, Armin: *Knowledge Discovery in Databases*. S. 51–108. In: *Knowledge Discovery in Databases: Eine Analyse des Änderungsmanagements in der Produktentwicklung*. Wiesbaden : Springer Fachmedien Wiesbaden, 2013. – http://dx.doi.org/10.1007/978-3-658-02002-6_3. – ISBN 978-3-658-02002-6