

---

## Kurzfassung

Das Aroma-Matching kann als Versuch aufgefasst werden, ein Zielaroma zu reproduzieren. Zu diesem Zweck werden Aromarezepturen durch instrumentelle Analysetechniken ausgewertet und von einem Aromenentwickler so lange verfeinert, bis eine sensorische Übereinstimmung erreicht oder das entwickelte Aroma vom Kunden akzeptiert wird. Dieser Trial-and-Error-Prozess ist zeitaufwändig und kostspielig, weil mehrere Versuchsrezepturen entwickelt werden müssen. Zusätzlich zum hohen Rohstoffverbrauch, können Rezepturen entstehen, die zu komplex oder zu teuer sind, um sie zu kommerzialisieren. Um das Matching zu erleichtern und die Entwicklungszeit eines Aromas zu verkürzen, ist eine datengetriebene Entscheidungsunterstützung notwendig. Eine Aromarezeptur setzt sich aus bis zu 50 Rohstoffen mit unterschiedlichen Anteilen zusammen. Deshalb wird das Aroma-Matching als multivariates Regressionsproblem aufgefasst und ein Empfehlungssystem vorgeschlagen, das automatisiert erste Abschätzungen der wichtigen Aromenbestandteile liefert und den Aromenentwickler mit initialen Rezepturvorschlägen unterstützt. Der zugrunde liegende, hochdimensionale Datensatz besteht aus gaschromatographischen Analysedaten der Zielaromen und Rezepturdaten der entstandenen Matches. Der Fokus dieser Arbeit liegt auf der Kombination von Dimensionsreduktions- und multivariaten Regressionsverfahren. Die PLS-Regression weist den niedrigsten kreuzvalidierten *average Relative Root Mean Squared Error* auf und eignet sich, um zukünftig die Rohstoffanteile eines Zielaromas vorherzusagen.

**Schlagerwörter:** Aroma-Matching, multivariate Regression, Dimensionsreduktion, Regularisierungsverfahren, Tobit-Regression.

---

## Abstract

Flavor matching can be considered as an attempt to reproduce a specific flavor. For this purpose, flavor formulations are analyzed by instrumental techniques and refined by a flavor developer until a sensory match is achieved or the developed flavor is accepted by the customer. This trial-and-error process is time consuming and costly because multiple trial recipes need to be developed. In addition to a high raw material consumption, flavor mixtures may be created that are too complex or too expensive to be commercialized. That's why there is a need for data-driven decision support in order to facilitate the initial estimates of important compounds. A flavor recipe consists of up to 50 raw materials with different proportions. Consequently, matching can be defined as a multi-target regression problem. For the described reasons, a recommendation system is proposed which automatically provides initial estimates of the important flavour components and aid the flavorist with recipe proposals. The underlying, high-dimensional data is composed of gas chromatographic data of the target flavors and recipe data of the resulting matches. This thesis focuses on dimension reduction and multivariate regression methods. The partial least square regression showed the lowest cross-validated *average Relative Root Mean Squared Error* and could be used to automatically generate recipe proposals in the future.

**Keywords:** flavour matching, multi-target regression, dimensionality reduction, regularization methods, tobit regression.