

Herkunftsunterscheidung von Weinen der Sorte 'Grüner Veltliner' anhand ihrer Aromaprofile mit Hilfe von Data Mining-Techniken und Neuronalen Netzwerken

ANDREAS SCHREINER¹, WALTER BRANDES¹, GIUSEPPE VERSINI², EMMERICH BERGHOFER³ und REINHARD EDER¹

¹Höhere Bundeslehranstalt und Bundesamt für Wein- und Obstbau
A-3400 Klosterneuburg Wiener Straße 74

²Istituto Agrario San Michele all'Adige
I-38010 San Michele all'Adige (TN), via Edmondo Mach, 1

³Universität für Bodenkultur
Department für Lebensmittelwissenschaften und -technologie
A-1190 Wien, Muthgasse 18

Im Rahmen vorliegender Arbeit wurde getestet, inwieweit anhand von Aromaanalysen bei österreichischen Weinen der Sorte 'Grüner Veltliner' eine korrekte Zuordnung der Herkunft (Weinbaugebiet) möglich ist. Die Aromaanalyse bestand aus einer Festphasenextraktion zur Trennung und Konzentrierung der freien und glykosidisch gebundenen Komponenten und anschließender Bestimmung mittels Gaschromatographie und Flammenionisationsdetektor. Die Konzentrationen von 176 Komponenten konnten bestimmt und die Aromasubstanzen zum Teil identifiziert werden. Zur Herkunftserkennung wurden 59 Weine der Sorte 'Grüner Veltliner', Jahrgang 2004, aus der weinbaulichen Praxis analysiert. Die Ergebnisse wurden mit Hilfe von statistischen und Data Mining-Techniken untersucht. Die Überprüfung der Tauglichkeit der einzelnen Methoden wurde über eine probeweise Vorhersage aller Proben mit Hilfe der leave one out-Methode durchgeführt. An statistischen Verfahren wurde der k-nächste Nachbarn (knn) Vergleich, die Lineare Diskriminanzanalyse (LDA) und ein Neuronales Netzwerk eingesetzt. Je nach Methode und Vorbereitung des Datensatzes konnten Trefferquoten zwischen 40,7 und 81,8% erzielt werden. Dabei zeigte sich, dass Weine aus Weinbaugebieten, von denen nur eine geringe Probenanzahl verfügbar war, deutlich schlechter vorhergesagt wurden. Die am besten geeignete Vorhersagemethode war das Neuronale Netzwerke mit einer Trefferquote von über 80%. Für den praktischen Einsatz der Methode ist es jedoch erforderlich, genügend Proben aus den zur Auswahl stehenden Gebieten zu analysieren, um damit Vergleichswerte und Trainingsdaten für die Zuordnung unbekannter Proben zu erstellen.

Schlagwörter: Wein, 'Grüner Veltliner', Aromaanalyse, Herkunftsbestimmung, Data Mining, Lineare Diskriminanzanalyse, Neuronale Netzwerke

Differentiation of origins of 'Grüner Veltliner' wines by means of aroma profiles and data mining techniques and neural networks. In this study, aroma analysis was carried out with Austrian wines of the cultivar 'Grüner Veltliner' with the intention to predict their origin (i.e. the specific growing area in Austria). The principle of the analysis was a solid phase extraction of the volatile compounds of a given wine sample, followed by separation and concentration of the extracted free and glycosidically bound components by means of GC and FID. The concentrations of a total of 176 components were measured and their aroma substances were partially identified. The results were examined by statistical and data mining techniques. The leave one out-technique was used to check the efficiency of the particular methods. The statistical methods used were the k-Next Neighbour comparison (knn), the Linear Discriminant Analysis (LDA) and a Neural Network. Depending on the method and the preparation of the datasets, hit ratios

ranging from 40.7 to 81.8% were reached. It was shown that a correct prediction of origin was much more unlikely for wines originating from a growing region from which only a few samples have been analysed. The application of a Neural Network emerged as the most suitable of all examined prediction methods (hit ratios up to about 80%). For a possible deployment of the method into day to day analytical practice it would be advisable to ensure the availability of an adequate number of samples from all winegrowing regions in question. This is necessary to provide comparison and training data for the prediction of unknown samples.

Key words: Wine, 'Grüner Veltliner', aroma analysis, determination of origin, Data Mining, linear discriminative analysis, neural networks

La distinction des origines de vins du cépage 'Grüner Veltliner' sur la base de leurs profils aromatiques à l'aide des techniques d'extraction des données (data mining) et des réseaux neuronaux. Dans le cadre du présent travail, il a été testé dans quelle mesure une attribution correcte de l'origine (région viticole) au moyen d'analyses des arômes est possible pour les vins autrichiens du cépage 'Grüner Veltliner'. L'analyse des arômes consistait en une extraction en phase solide destinée à la séparation et à la concentration des composants libres et de ceux faisant l'objet d'une liaison glycosidique, suivi de la détermination à l'aide de la chromatographie en phase gazeuse et du détecteur à ionisation de flamme. Il a été possible de déterminer les concentrations de 176 composants et d'identifier partiellement les substances aromatiques. 59 vins du cépage 'Grüner Veltliner', millésime 2004, issus de la pratique viticole, ont été analysés. Les résultats ont été examinés par des moyens statistiques et des techniques d'extraction des données (data mining). La vérification de l'aptitude des différentes méthodes a été effectuée en réalisant des pronostics de l'ensemble des échantillons à titre d'essai à l'aide de la méthode leave one out. Les procédures statistiques utilisées ont été la comparaison des k plus proches voisins (kppv), l'analyse discriminante linéaire (ADL) et un réseau neuronal. Des taux de pertinence entre 40,7 et 81,8 % ont pu être obtenus en fonction de la méthode et de la préparation des enregistrements. Dans ce contexte, il s'est révélé que les pronostics pour des vins en provenance de régions viticoles, pour lesquelles on ne disposait que d'un faible nombre d'échantillons, avaient été beaucoup moins corrects. La méthode de pronostic la plus appropriée avec un taux de pertinence de plus de 80 % a été celle des réseaux neuronaux. Pour la mise en pratique de cette méthode, il est cependant nécessaire d'analyser une quantité suffisante d'échantillons provenant des régions à sélectionner, afin d'établir des valeurs de comparaison et des données d'apprentissage pour le classement d'échantillons inconnus.

Mots clés : vin, 'Grüner Veltliner', analyse des arômes, détermination de l'origine, data mining, analyse discriminante linéaire, réseaux neuronaux