

Wie gelangen wir zu den Messwerten, Daten und dem Match Score:

Zunächst wird eine Probe des kosmetischen Produktes auf ein Papier aufgetragen, um die Farbinformationen zu messen („die Probe“). Jede Probe wird in demselben Gerät, welches für die Analyse des Erscheinungsbildes von festen Texturen entwickelt wurde, gemessen. Für jede einzelne Probe werden die folgenden Messwerte aufgezeichnet:

1 **Messungen**

a) **Farbmessungen („Color“)**

2 Das Gerät misst zunächst die Farbe der Probe. Dazu wird der spektrale Reflektionsgrad der Substanz unter verschiedenen Bedingungen gemessen. Der Reflektionsgrad ist der Anteil des auftreffenden Lichts, der von einer Fläche reflektiert wird (siehe hierzu etwa <https://de.wikipedia.org/wiki/Reflexionsgrad> bzw. <https://en.wikipedia.org/wiki/Reflectance>). Die unter den verschiedenen Bedingungen gemessenen Spektraldaten werden dann unter Zuhilfenahme der in DIN EN ISO 11664 beschriebenen Farbanpassungsfunktion in den sogenannten „CIE-Lab-Raum“ übertragen. Die DIN EN ISO 11664 legt Spektralwertfunktionen zur Anwendung in der Farbmeterik fest. Die jeweiligen Dokumente sind abrufbar unter <https://www.din.de/de>. Das CIE-Lab-System ist ein Farbraum, der von der internationalen Beleuchtungskommission CIE im Jahr 1976 festgelegt wurde. Das CIE-Lab System ist das heute wohl gebräuchlichste Farbsystem und beschreibt alle wahrnehmbaren Farben (vgl. hierzu im Einzelnen <https://www.cielab.de/>, <http://cie.co.at/> sowie <https://de.wikipedia.org/wiki/Lab-Farbraum>).

b) **Glanzmessungen („Gloss“)**

3 Das Gerät misst den Glanz („Gloss“) der Probe. Die Glanzmessung stellt das Verhältnis zwischen reflektiertem Licht und einfallendem Licht für die Probe im Vergleich zum Verhältnis zwischen reflektiertem Licht und einfallendem Licht für

den Referenzwert dar. Dies entspricht der DIN EN ISO 2813 zur Bestimmung des Glanzwertes und wird dort im Einzelnen beschrieben.

c) **Glitzermessungen („Glitter“)**

- 4 Das Gerät misst den Glitzeranteil („Glitter“) der Probe. Hierbei wird unter einer Kamera der Effekt der Glitzerpartikel unter verschiedenen Wetterbedingungen simuliert und gemessen. Die Simulation entspricht der wissenschaftlichen Forschung nach *Alman, D.H., "Directional Color Measurement of Metallic Flake Finishes" (1987).*

2. **Vergleich**

- 5 Die Produkte werden segmentiert und in die verschiedenen Kategorien (Lippenstift, Lipgloss, Nagellack) eingeteilt. Die ermittelten Messwerte werden innerhalb der jeweiligen Kategorie verglichen. Dazu werden alle Messwerte aus derselben Kategorie einer sogenannten Min-Max-Skalierung unterworfen. Die Skalierung richtet sich nach dem Datenanalyse-Tool „scikit-learn“ (abrufbar unter <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>). Hierbei werden alle Messwerte in den Wertebereich von 0 bis 1 überführt. Dadurch wird sichergestellt, dass die Daten gleichwertig in das Ergebnis einfließen.

Berechnung von vier Eckpunkten für jedes Segment:

1. Δ_{gesamt} : Die größte euklidische Distanz, die wir zwischen allen Produkten in diesem Segment zueinander finden können auf Basis aller Merkmale.
2. Δ_{Farbe} : Die größte euklidische Distanz, die wir zwischen allen Produkten in diesem Segment zueinander finden können auf Basis der Merkmale die Farbeigenschaften beschreiben.
3. Δ_{Glitzer} : Die größte euklidische Distanz, die wir zwischen allen Produkten in diesem Segment zueinander finden können auf Basis der Merkmale die Glitzereigenschaften beschreiben.

4. Δ_{Glanz} : Die größte euklidische Distanz, die wir zwischen allen Produkten in diesem Segment zueinander finden können auf Basis der Merkmale die Glanzeigenschaften beschreiben.

Berechnung für jedes Produkt von vier Distanzen zu jedem anderen Produkt in dem Segment:

1. δ_{gesamt} : Die euklidische Distanz auf Basis aller Merkmale.
2. δ_{Farbe} : Die euklidische Distanz auf Basis der Merkmale die Farbeigenschaften beschreiben.
3. δ_{Glitzer} : Die euklidische Distanz auf Basis der Merkmale die Glitzereigenschaften beschreiben.
4. δ_{Glanz} : Die euklidische Distanz auf Basis der Merkmale die Glanzeigenschaften beschreiben.

Normalisierung

Diese Distanzen normalisieren wir mit den entsprechenden Eckpunkten (Strecke zwischen 0 und dem unähnlichsten Produkt) und erhalten folgende normalisierten Distanzen:

$$\text{normalisiert } \delta_{\text{gesamt}} = \delta_{\text{gesamt}} / \Delta_{\text{gesamt}}$$

$$\text{normalisiert } \delta_{\text{Farbe}} = \delta_{\text{Farbe}} / \Delta_{\text{Farbe}}$$

$$\text{normalisiert } \delta_{\text{Glitzer}} = \delta_{\text{Glitzer}} / \Delta_{\text{Glitzer}}$$

$$\text{normalisiert } \delta_{\text{Glanz}} = \delta_{\text{Glanz}} / \Delta_{\text{Glanz}}$$

Diese Distanzen liegen immer im Wertebereich 0 bis 1. Je höher eine Distanz, desto unterschiedlicher sind die Produkte. Um den Matchscore zu erhalten, also einen Wert zu bekommen, der umso größer ist, je ähnlicher sich Produkte, berechnen wir die Differenz von 1. Durch mathematische Rundung und der Umwandlung in einen Prozentwert erhält man den Matchscore für die verglichenen Produkte.

3. Berechnung des „Match Score“

6 Zur Berechnung des „“ werden alle n Produkte $\{i\}_{1 \leq i \leq n}$ und die gewichteten h Messwerte $\vec{m}_i := \{m_{i,k}\}_{1 \leq k \leq h}$ aus der gleichen Kategorie miteinander verglichen, indem der gewichtet euklidische Abstand zwischen allen skalierten Messwertpaaren berechnet wird (Messwert 1 von Produkt X mit Messwert 1 von Produkt Y, Messwert 2 von Produkt X mit Messwert 2 von Produkt Y usw.). Daraus ergibt sich eine Matrix mit folgenden Abständen: Produkt 1 mit Produkt 1, Produkt 1 mit Produkt 2, Produkt 2 mit Produkt 1, Produkt 2 mit Produkt 2, usw.

$$\delta = \{\delta_{i,j}\}, \quad \delta_{i,j} = |\vec{w}(\vec{m}_i - \vec{m}_j)|$$

7 Die berechnete Distanzmatrix enthält alle Abstände zwischen allen möglichen Produktpaaren, einschließlich sich selbst. Gleichzeitig wird aus allen möglichen Messwerten aller Produkte das Element mit dem höchsten Zahlenwert ermittelt und gespeichert.

$$\Delta = \max(m_{i,k} | 1 \leq i \leq n, 1 \leq k \leq h)$$

8 Für jedes Produkt werden dann alle möglichen Distanzen in aufsteigender Reihenfolge sortiert (kleinste zuerst) und die besten Übereinstimmungen angezeigt. Um die prozentuale Genauigkeit zu berechnen, wird die Entfernung für jede der besten Übereinstimmungen durch das numerisch höchste Element dividiert und von 1 subtrahiert.

$$1 - \frac{\delta}{\Delta}$$

9 Die gleiche Berechnung wird für $m_{i,k}$ für die „Farbmessungen“, die „Glanzmessungen“ und die „Glitzermessungen“ durchgeführt. Das Ergebnis zeigt alle Farbabstände, alle Glanzabstände und alle Unterschiede in der Glitzerpartikeldichte und -Größe zwischen allen möglichen Produktpaaren.

10 Dabei werden für die einzelnen Parameter Farbe (Color), Glanz (Gloss) und Glitzer (Glitter) unterschiedliche Ergebnisse ermittelt. Diese werden

unterschiedlich gewichtet, wobei die Messwerte für die Farbe mit etwa 78% das Ergebnis bestimmen, die Messwerte für den Faktor Glanz mit etwa 18% und die Messwerte für Glitter mit 4% einfließen. Deshalb weicht das Gesamtergebnis auch in der Regel vom Durchschnittsergebnis der 3 Parameter ab. Insgesamt kommt so der beste Match zu Stande. Ein Match wird erst ab 80 % Farbübereinstimmung angezeigt. Eine 100% Farbübereinstimmung gibt es in der Regel im „DupeFinder“ nicht.