

Dublettscore: Kennzahl zur Qualität von Stammdaten im Einkauf

Der Dublettscore ist eine Kennzahl, die misst, wie weit Stammdatenbestände – etwa Material-, Lieferanten- oder Artikelstämme – frei von Doppel-einträgen sind. Er wird typischerweise als prozentualer Wert ausgedrückt und ist ein zentraler Indikator für die Datenqualität im Stammdatenmanagement. Für Einkäufer ist er relevant, weil Dubletten zu doppelten Beständen, falschen Preisvergleichen und verfälschten Auswertungen führen können.

Das Wichtigste in Kürze

- Der Dublettscore setzt die Anzahl eindeutiger Datensätze ins Verhältnis zur Gesamtzahl der Datensätze.
- In der Praxis liegt ein guter Wert häufig bei über 95 Prozent; darunter steigt der Aufbereitungsaufwand deutlich.
- Dubletten verzerren Spend-Analysen, Preisvergleiche und die Purchase Price Variance (PPV).
- Die Ermittlung erfolgt in der Regel über fuzzy matching, also unscharfe Ähnlichkeitssuche in Namen, Nummern und Attributen.
- Ein niedriger Score erhöht typischerweise den Prüfaufwand im Rechnungsfreigabe-Workflow.
- Regelmäßige Bereinigung ist Teil eines funktionierenden Stammdatenmanagement und reduziert Doppelbestellungen.

Definition: Dublettscore

Der Dublettscore quantifiziert den Anteil doppelter oder potenziell doppelter Datensätze an einem Stammdatenbestand. Er dient als Messgröße für Datenqualität und als Steuerungsgröße für Datenbereinigungsprojekte.

Berechnungsgrundlage

In der Praxis wird der Score häufig als Anteil der als eindeutig klassifizierten Datensätze an allen Datensätzen berechnet. Alternativ zählen Unternehmen die identifizierten Dublettenkandidaten und setzen sie ins Verhältnis zum Gesamtbestand. Entscheidend ist, dass die Definition unternehmensintern einheitlich dokumentiert wird, damit Werte über Zeiträume vergleichbar bleiben.

Abgrenzung zur Fehlerquote

Der Dublettscore betrachtet nur eine Qualitätsdimension: Redundanz. Andere Fehler wie unvollständige Attribute, falsche Maßeinheiten oder veraltete Lieferantenzuordnungen werden separat erfasst. Für ein ganzheitliches Stammdatenmanagement kombiniert man den Score daher mit weiteren Datenqualitätskennzahlen.

Methoden und Vorgehen zur Dublettenermittlung

Die Ermittlung des Dublettenscores folgt einem mehrstufigen Prozess: Profilierung, Matching, Klassifizierung und Bereinigung. Der Score wird dabei vor und nach der Bereinigung gemessen, um den Fortschritt nachzuweisen.

Matching-Verfahren

Exakte matches über Nummern reichen selten aus, da Dubletten oft durch Tippfehler, abweichende Bezeichnungen oder unterschiedliche Maßeinheiten entstehen. Üblich ist daher fuzzy matching mit Ähnlichkeitswerten, ergänzt um Match-&-Merge-Regeln, die festlegen, ab welchem Ähnlichkeitsgrad zwei Datensätze als Dublette gelten. Attribute wie Maße, Werkstoffe und Normen dienen als Bestätigungskriterien.

Klassifizierung und Freigabe

Gefundene Kandidaten werden in der Praxis selten automatisch zusammengeführt. Häufig durchläuft ein Treffer eine manuelle Prüfung durch Fachbereiche, ähnlich wie eine Nachforderung von Unterlagen im Lieferantenmanagement. Erst nach Freigabe werden Datensätze zusammengeführt, wobei Historien wie Bestellungen und Preise erhalten bleiben müssen.

Wichtige KPIs rund um den Dublettenscore

Der Dublettenscore steht in der Praxis nie allein. Er wird mit ergänzenden Kennzahlen kombiniert, um Aussagen über Kosten, Prozessqualität und Datenverlässlichkeit treffen zu können.

Datenqualitäts-Kennzahlen

Neben dem Score zählen Vollständigkeitsgrade, Aktualität der Stammdaten und die Anzahl neu angelegter Datensätze pro Periode. Eine steigende Neuanlagenquote bei sinkendem Score deutet häufig darauf hin, dass Suchfunktionen oder Anlageprozesse unzureichend sind.

Auswirkung auf Einkaufskennzahlen

Dubletten verzerren Spend-Bündelungen und damit potenzielle Skaleneffekte bei der Lieferantenkonsolidierung. Auch die Purchase Price Variance (PPV) kann verzerrt erscheinen, wenn Preise auf unterschiedliche Datensätze desselben Materials gebucht werden. Der Score ist damit auch ein Belastbarkeitstest für Kennzahlen im Businessplan Einkauf.

Risiken und Kontrollen

Ein unbemerkter schlechter Dublettenscore wirkt sich entlang der gesamten Beschaffungskette aus. Kontrollen sollten daher prozessintegriert statt einmalig erfolgen.

Operative Risiken

Typische Folgen sind doppelte Bestände und Bestellungen, fehlerhafte Rahmenverträge und Mehrfachzahlungen an denselben Lieferanten unter leicht abweichenden Namen. Im Rechnungsfreigabe-Workflow steigt der manuelle Klärungsaufwand, wenn Rechnungen auf unterschiedliche Datensätze desselben Vorgangs treffen.

Strategische Risiken

Auf strategischer Ebene verzerren Dubletten das Lieferantenportfolio: Ersatzteile werden als Einzelteile beschafft statt gebündelt, und Lieferantenbewertungen – etwa im Rahmen der Pre-Qualification oder beim Lieferantenzertifikat – basieren auf unvollständigen Daten. Auch für Revisionen und Exportkontrolle sind saubere Stammdaten Voraussetzung.

Kontrollmaßnahmen

Bewährt haben sich Duplikatprüfungen bereits bei der Neuanlage, periodische Score-Messungen mit Trendverfolgung und klar definierte Verantwortlichkeiten im Stammdatenmanagement. Zusätzlich regeln Einkaufsrichtlinien, etwa eine Einkaufsrichtlinie MRO vs. Capex, wer Materialstämme anlegen darf und nach welchen Kriterien geprüft wird.

Aktuelle Entwicklungen: Automatisierung und Datenpflege

Durch KI-gestützte Klassifikation und semantische Suche verbessern sich die Erkennungsraten bei Dubletten deutlich, insbesondere bei Freitextartikelnummern. Gleichzeitig sinkt der manuelle Aufwand für Prüfung und Zusammenführung.

KI und semantisches Matching

Moderne Verfahren erkennen Dubletten auch bei stark abweichenden Bezeichnungen, etwa zwischen Hersteller- und Werkbezeichnungen. In der Praxis werden Vorschläge der Software dennoch meist durch Fachabteilungen bestätigt, um Fehl-Merges zu vermeiden, die schwer rückgängig zu machen sind.

Verzahnung mit Einkaufsprozessen

Sauber aufbereitete Stammdaten sind Voraussetzung für automatisierte Prozesse wie Available to Promise (ATP) in der Verfügbarkeitsprüfung oder konsistente Preise inklusive Raw Material Surcharge. Auch im Ramp-up-Management neuer Produkte zahlt ein hoher Dublettenscore direkt auf Termintreue ein, weil Ersatzteil- und Materialdaten zuverlässig zuordenbar sind.

Typische Kennzahlen

Dublettenscore: > 95 %

typischer Richtwert für einen solide bereinigten Stammdatenbestand

Neuanlagenquote mit Dublette: < 2 %

Anteil neuer Datensätze, die später als Dublette identifiziert werden

Bereinigungsquote pro Quartal: > 90 %

Anteil bestätigter Dublettenkandidaten, die tatsächlich zusammengeführt werden

Praxisbeispiel

Ein mittelständischer Maschinenbauer mit rund 120.000 Materialstammsätzen stellte bei einer Spend-Analyse fest, dass mehrere Normteile unter unterschiedlichen Bezeichnungen geführt wurden. Der anfängliche Dublettenscore lag im Bereich um 92 Prozent eindeutiger Datensätze. Nach einer Bereinigung mit fuzzy matching und Review durch die Fachabteilung stabilisierte sich der Wert über mehrere Quartale im Bereich über 97 Prozent.

- Einführung einer Pflichtsuche vor Neuanlage von Materialstammsätzen

- Definition unternehmensweiter Match-&-Merge-Regeln mit Freigabestufen
- Quartalsweise Score-Messung mit Bericht an Einkauf und IT
- Bündelung von Normteilen auf fewer Lieferanten zur Hebung von Skaleneffekten

Fazit

Der Dublettscore ist eine einfache, aber wirkungsvolle Kennzahl für die Qualität von Einkaufs- und Materialstammdaten. Er macht Redundanz sichtbar, die sonst verdeckt Kosten erzeugt – durch Doppelbestellungen, verzerrte Auswertungen und erhöhten Prüfaufwand. Wichtig ist eine einheitliche Berechnungsmethode und eine regelmäßige Messung, nicht ein einmaliger Bereinigungslauf. In Kombination mit klaren Anlageprozessen und definierten Verantwortlichkeiten trägt der Score unmittelbar zur Einkaufsoptimierung bei. Für technische Einkäufer ist er damit ein verlässlicher Frühindikator für die Belastbarkeit sämtlicher datenbasierter Entscheidungen.

Häufige Fragen

Was ist ein Dublettscore?

Der Dublettscore ist eine Kennzahl für den Anteil doppelter Datensätze in einem Stammdatenbestand, etwa im Material- oder Lieferantenstamm. Er wird typischerweise als Prozentwert ausgedrückt. Ein hoher Wert bedeutet eine geringe Anzahl von Dubletten und damit eine gute Datenqualität.

Wie berechnet man den Dublettscore?

Üblich ist die Formel: eindeutige Datensätze geteilt durch Gesamtanzahl der Datensätze, multipliziert mit 100. Die der Dubletten erfolgt häufig über fuzzy matching und definierte Match-&-Merge-Regeln. Wichtig ist, die Methodik zu dokumentieren, damit Werte über Perioden vergleichbar bleiben.

Ab wann ist ein Dublettscore gut?

In der Praxis gelten Werte über 95 Prozent als solide, über 98 Prozent als sehr gut. Der Richtwert hängt jedoch von Branche und Datenbestand ab. Entscheidender als der Absolutwert ist der Trend: Ein sinkender Score weist auf Probleme im Anlageprozess hin.

Welche Kosten verursachen Dubletten im Einkauf?

Dubletten führen zu doppelten Beständen, verpassten Bündelungseffekten und Mehraufwand im Rechnungsfreigabe-Workflow. Zudem verzerren sie Spend-Analysen und die Purchase Price Variance (PPV). Konkrete Höhe der Kosten variiert je nach Unternehmen und lässt sich nur im Einzelfall ermitteln.

Wie senkt man Dubletten in Stammdaten dauerhaft?

Entscheidend ist Prävention: Pflichtsuche vor Neuanlage, standardisierte Benennungskonventionen und definierte Anlageberechtigungen. Ergänzend helfen periodische Dublettenläufe mit klarer Freigabeprozessur. Einmalige Großbereinigungen ohne Prozessänderung führen in der Praxis häufig zu einem erneuten Anstieg.