



**Fachbereich
Humanwissenschaften
Institut für Cognitive Science**

Bachelorarbeit

Automatisierte Analyse von Preisbildungsmustern
anhand von Zeitreihendaten der Markttransparenzstelle für Kraftstoffe

Autor: Kai Fritsch
kai.m.fritsch@gmail.com

Prüfer: Prof. Dr. Oliver Vornberger

Inhaltsverzeichnis

1	Einleitung	1
1.1	Geschichtlicher Hintergrund	1
1.2	Pricing Strategien	1
1.3	Motivation	2
1.3.1	Unmittelbar Nutzen	2
1.3.2	Potentielle Möglichkeiten	3
2	Grundlagen	4
2.1	Datensatz	4
2.2	PSQL Datenbank	6
2.2.1	Funktion	6
2.2.2	Indexing	6
2.3	Verwendete Programme und Packages	6
2.4	Pricing-Tool	6
2.4.1	Funktionsweise	6
2.4.2	Konkurrenz bestimmen	7
2.4.3	Regeln festlegen	7
2.4.4	Automatisierung wählen	8
3	Anforderungsanalyse	9
3.1	Zielsetzung	9
3.2	Generelle Voranalyse	9
3.2.1	Datenmenge	10
3.2.2	Strukturierung des Datensatzes	10
3.2.3	Algorithmen und andere Ansätze	11
3.3	Modularisierung	13
4	Umsetzung	14
4.1	Datenaufbereitung	14
4.2	Tankstellendaten	14
4.2.1	Preisdaten	14
4.3	Potentielle Konkurrenten ermitteln	14
4.4	Reaktionen erkennen	15
4.4.1	Zeitliche Komponente	16
4.4.2	Semantische Komponente	17
4.4.3	Fallunterscheidung	17
4.5	Regeln spezifizieren	20
4.5.1	Maximaler Preisabstand	20
4.5.2	Zeitintervall	22
4.5.3	Klassifikation Konkurrenz	22
4.6	Automatisierungsgrad bestimmen	22
4.6.1	Formen der Automatisierung	23
4.6.2	Regelverletzung und Reaktionszeiten	23
5	Ergebnisse	24

5.1	Beschreibung	24
5.2	Diskussion	24
5.2.1	Statistiken	24
5.2.2	Problemanalyse	24
5.3	Zusammenfassung	24
6	Fazit	25
6.1	Weitere Möglichkeiten	25
6.1.1	Gegenrichtung prüfen	25
6.1.2	Reaktionen an Regeln trimmen	25
6.1.3	Auslöser untereinander vergleichen	25
6.1.4	Constraint Satisfaction	25

1 Einleitung

1.1 Geschichtlicher Hintergrund

Kraftstoffpreise unterliegen nun schon seit längerer Zeit großen Preisschwankungen. Die hohe Homogenität und sinkende Nachfrage sind optimale Bedingungen für starken Konkurrenzdruck. Der Druck ist so hoch, dass die Anzahl an Tankstellen in Deutschland seit 1970 von gut 46.000 auf heute knapp 15.000 gefallen ist.[?] Nichts desto trotz ist die Anzahl an Preisänderungen in den letzten Jahren noch einmal merklich gestiegen. Während der Preis 2011 noch durchschnittlich weniger als zwei mal pro Tag geändert wurde, sind die meisten Tankstellen inzwischen bei um die sechs Änderungen pro Tag angelangt.[?]

Mitgrund für diese jüngste Steigerung könnten diesmal jedoch die neuerlichen elektronischen Möglichkeiten sein. Während Tankstelleninhaber vor einigen Jahren noch ihrer Konkurrenz einen Besuch abstatten mussten um deren Preise auszukundschaften, reicht heute eine Suchanfrage im Internet. Und selbst diese ist seit der Einführung der Markttransparenzstelle für Kraftstoffe nicht mehr notwendig. Jede Tankstelle ist seit Ende 2013 verpflichtet ihre Preisänderungen dem Bundeskartellamt für Kraftstoffe elektronisch zu übermitteln. Der auf diese Weise entstandene Datensatz bildet die Grundlagen dieser Arbeit.

1.2 Pricing Strategien

Tankstellen waren in den letzten Jahren immer wieder gezwungen ihre Strategien dem immer schneller werdenden Markt anzupassen. Die mit der neuerlichen Änderung verbundene zentrale und unmittelbare Verfügbarkeit aller aktuellen Preise ermöglicht es Tankstellenbesitzern ihre Preise innerhalb kürzester Zeit an die Konkurrenz anzupassen. Das bestehen einer einheitlich definierten Schnittstelle ermöglicht es auch Preisänderungen zunächst automatisch zu registrieren und anschließend sogar die diesbezüglich relevanten eigenen Preise anhand von Regeln ebenfalls automatisch anpassen zu lassen. So eine Automatisierung birgt aber auch Probleme.

Die Vereinfachung der Arbeit durch automatische Prozesse führt auch dazu, dass mehr Konkurrenten bei gleichzeitig geringerem Aufwand überwacht werden können und die jeweilig gewünschten Reaktionen viel zuverlässiger durchgeführt werden können. Somit steigt aber auch die generelle Anzahl der Preissenkungen und damit auch die eigenen Preise. Um dem entgegen zu wirken müssen die Preise zu Anfang des Tages höher angesetzt werden. Inzwischen wurde sogar eine Mittagserhöhung eingeführt um dem Preisverfall

entgegen zu wirken. Auch eine gewisse Abstimmung von Preisen zwischen den Konkurrenten wird immer wichtiger. Wollen zwei Tankstellen jeweils einen günstigeren Preis fahren als die jeweils andere, so würde der Preis kontinuierlich sinken, bis eine Tankstelle ihr Minimum erreicht hat. Es gibt unter Tankstellen deshalb so etwas, wie eine Rangordnung, die den Markt in A-Preiser und B-Preiser unterteilt. A-Preiser, zumeist die Markentankstellen, erlauben den B-Preisern, zumeist freie Tankstellen oder kleinere Ketten, etwas geringer Preise zu fahren. Während der Preisverfall bei einem Verstoß gegen diese Rangordnung vor der Automatisierung noch relativ langsam von statten ging, liegt der Preis bei automatischem Pricing innerhalb weniger Minuten beim Minimum. Zudem muss jegliche Art von Sonderfällen in den Regeln berücksichtigt werden. Um ungewünschten Preisen durch einen vollautomatischen Betrieb vorzubeugen beschränkt sich die Automatisierung teilweise auf den Vorschlag von Preisänderungen, welche dann noch manuell durchgeführt werden müssen.

1.3 Motivation

Die Idee dieser Arbeit besteht darin das Konkurrenzverhalten der Tankstellen zu analysieren. Es ist anzunehmen, dass eine Tankstelle egal ob manuell oder automatisch ihr Preise nicht zufällig anpasst, sondern versucht sie möglichst gewinnoptimierend zu gestalten. Preise sollten also generell nach bestimmten Richtlinien in Abhängigkeit von verschiedenen Parametern festgelegt werden. Wo konsistent nach bestimmten Parametern Entscheidungen getroffen werden, entstehen Regeln, die auch zurückverfolgt werden können. Je konsistenter eine Regel umgesetzt wird, desto einfacher ist es diese anhand ihrer Auswirkungen zu rekonstruieren. Je automatisierter also eine Tankstelle ihr Pricing betreiben, desto einfacher sollte es sein ihr Verhalten zu analysieren.

1.3.1 Unmittelbar Nutzen

In einem Wettbewerb ist es immer von Vorteil die Strategien der Konkurrenten zu kennen. Es ermöglicht eine akkuratere Vorhersage des zukünftigen Verlaufes und erlaubt es einem somit seine eigene Strategie gewinnbringender zu gestalten. Da das Pricing in großen Ketten zentral betrieben wird, liegt die Verantwortung über das Preisverhalten von mehreren Hunderten Tankstellen bei einem sehr kleinen Personenkreis. Bei beispielsweise 100 Tankstellen mit jeweils fünf potentiellen Konkurrenten und sechs Preisänderungen pro Tag müssten diese Personen Regeln für 500 Konkurrenten in 3000 Preisänderungen täglich erkennen. Eine automatisierte Regelerkennung würde also viel Arbeit und Mühen ersparen und wahrscheinlich sogar verlässlichere Ergebnisse liefern. Die eingesparte Zeit kann

also in die Optimierung des eigenen Verhaltens investiert werden und die verlässlicheren Ergebnisse liefern eine bessere Grundlage für gezielte Optimierungen.

1.3.2 Potentielle Möglichkeiten

Jede Preissenkung verringert die Gewinnspanne einer Tankstelle. Es liegt also prinzipiell nicht im Interesse einer Tankstelle Preise zu senken. Angenommen eine automatisierte Regelerkennung wäre gegeben. Es bestehen einige Möglichkeiten dieses Wissen gewinnbringend zu nutzen:

- Mit einer solchen Regelerkennung ließen sich reaktive Preisänderungen von aktiven unterscheiden. Einzelne Änderungen könnten somit iterativ über die jeweiligen auslösenden Konkurrenten zurückverfolgt werden bis hin zu einer aktiven Änderung, also dem eigentlichen Auslöser. Die Erkennung von Auslösern bietet dann Möglichkeiten andersweitig zu reagieren und Senkungen zu umgehen.
- Ebenso denkbar wäre, dass es diese Auslöser in Form von aktiven Senkungen gar nicht so häufig gibt. Angenommen Tankstellen würden zwar bei Senkungen aufeinander reagieren, ihre Erhöhungen aber unabhängig gestalten. Eine Tankstelle die ihre Preise weniger hoch erhöht wäre somit potentieller Auslöser von einer Senkungswelle. Es wäre also unter Umständen möglich, dass man sich intensiver mit den eigenen Erhöhungen beschäftigen müsste.
- Es ist bereits ohne dieses Tool möglich einige Preisänderungen über größere Strecken zurückzuverfolgen. Mit dem Tool bestünde die Möglichkeit Verkettungen von Konkurrenten zu erkennen und diese Wege exakt aufzuzeigen. Man könnte dann versuchen diese Ketten zu unterbrechen, sodass isoliertere Preissektoren entstünden.
- Es bietet auch die Möglichkeit die eigenen Änderungen systematisch zu analysieren und weniger wichtige Konkurrenten aus dem System zu nehmen.

Diese Änderungen müssen jedoch nicht zwangsläufig nur den Tankstellen zu Gute kommen. Eine Tankstelle ist gesetzlich verpflichtet keine verlustbringenden Preise zu fahren um Preisdumping zu vermeiden. Sie muss somit die ganzen Preissenkungen an einem Tag in die Preiserhöhung zu Tagesbeginn einkalkulieren. Das ist höchstwahrscheinlich auch ein Grund für die erst kürzlich eingeführte Mittagserhöhung. Eine Verminderung der Preissenkungen führt somit nicht zwangsweise zu höheren Preisen, da weiterhin ein Wettbewerb besteht. Es sollte vielmehr zu generell stabileren Preisen führen, sodass Kunden sich besser auf Preise verlassen können.

2 Grundlagen

Dieses Kapitel soll einen Überblick über die mir zur Verfügung stehenden Informationen, sowie die von mir verwendeten Tools liefern. Zu den Informationen zählen zum Einen der Datensatz der Markttransparenzstelle(MTS) für Kraftstoffe, sowie zum Anderen der exemplarische Aufbau eines der auf dem Markt verfügbaren automatischen Pricing-Tools. Zusätzlich dazu habe ich noch einen groben Erfahrungsbericht, wie dieses Tool bei einer Tankstelle eingesetzt wird.

2.1 Datensatz

„Seit dem 31. August 2013 sind Unternehmen, die öffentliche Tankstellen betreiben oder über die Preissetzungshoheit an diesen verfügen, verpflichtet, Preisänderungen bei den gängigen Kraftstoffsorten Super E5, Super E10 und Diesel in Echtzeit an die Markttransparenzstelle für Kraftstoffe zu melden.“[?]

In Echtzeit bedeutet, dass die Information über eine Preisänderung spätestens 5 Minuten nach der Änderung selber an der MTS eingehen müssen. Neben den Preisänderungen müssen auch die Stammdaten der Tankstellen wöchentlich an die MTS gemeldet werden. Die MTS veröffentlicht die Daten nicht direkt selber, sondern stellt sie Verbraucherinformationsdiensten(VID) zur Verfügung. Diese verpflichten sich die Daten in irgendeiner den Verbrauchern nützlichen Form zu veröffentlichen. Die Schnittstelle der MTS ist in der folgenden Abbildung dargestellt.

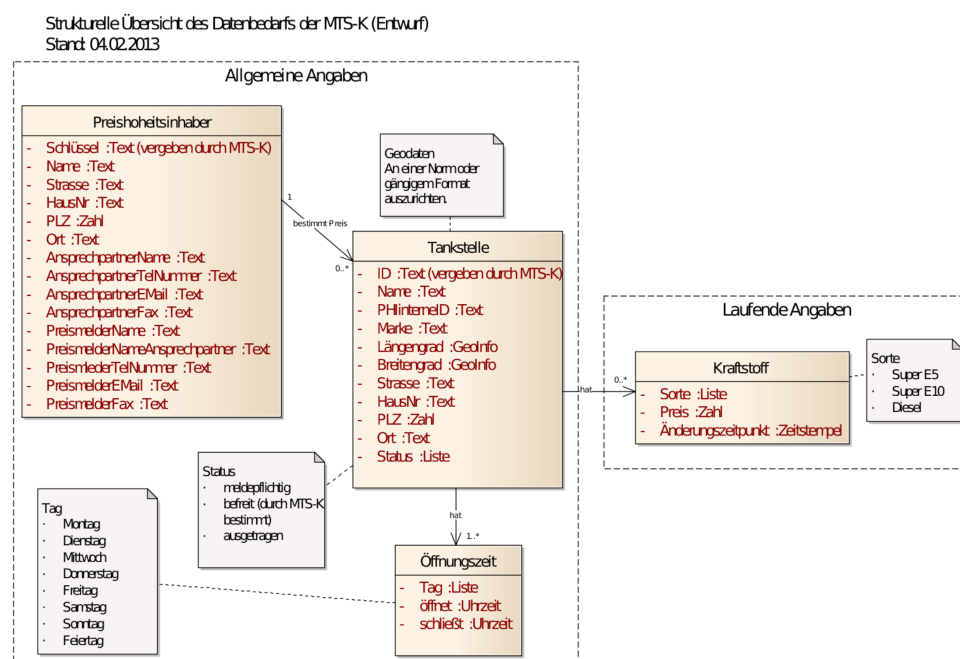


Abbildung 1: Datenstruktur der MTS-K[?]

Die Daten für diese Arbeit entstammen dem Verbraucherinformationsdienst der Tankerkönig API. Tankerkönig bezieht seine Daten direkt von der MTS und stellt diese sowohl im Zuge einer Echtzeit-Benzinpreis-API als auch gesondert aufbereitet als historischen Datensatz in Form eines PostgreSQL dumps im 9.4.-Format zur Verfügung. Die für diese Arbeit verwendeten dumps umfassen neben den kompletten Preisdaten vom 8.6.2014 bis heute auch die Informationen über die Stammdaten der Tankstellen.[?]

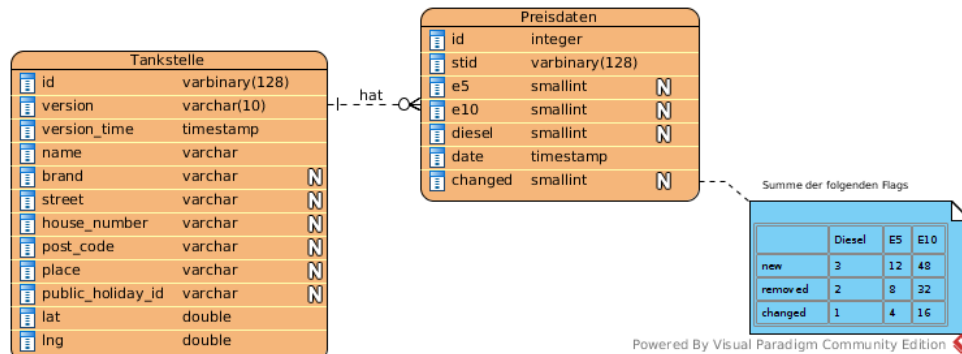


Abbildung 2: ERD des historischen Datensatzes

Alle für diese Arbeit verfügbaren Informationen sind in der Abb.2 abgebildet. Die Daten aus dem MTS-Datensatz werden größtenteils von den Tankstellen selber zur Verfügung gestellt, weshalb die Felder teilweise nicht einheitlich definiert sind. Die *id* wird als einziges von der MTS bereitgestellt. Die Felder *version_ttimestamp* und *version* wurden dem Datensatz von Tankerkönig hinzugefügt und identifizieren wann Änderungen an den Stammdaten einer Tankstelle vorgenommen wurden, beziehungsweise zählen die verschiedenen Versionen. Die Felder *name* und *brand* werden von den Tankstellen selber befüllt und enthalten sehr verschiedene Informationen. Bei großen Markentankstellen wird unter *brand* die Marke geführt und als *name* wird hier oft schon die Adresse aufgeführt, da diese offenbar als Identifizierung innerhalb des Unternehmens dient. Einzelne Tankstellen oder kleinere Unternehmen machen zu *brand* entweder keine Angaben oder irgendwie auf andere Weise kenntlich, dass sie keinem großen Konzern angehören. Bei *name* geben Sie dann meistens den Inhaber oder den Unternehmensnamen an. Bei den Feldern zur Adresse fehlen teilweise Informationen oder sie werden zusammen aufgeführt. Beispielsweise taucht die Hausnummer des öfters schon im Straßenfeld mit auf. Der public holiday identifier gibt an welchem Bundesland dieser Standort angehört und damit die entsprechend geltenden Feier- und Ferientage. Der Standort kann über *lat* und *lng* eindeutig identifiziert werden.

Bei den Preisänderungen hat Tankerkönig die neu angeschlagenen Preise der Kraftstoffsorten Diesel, E5 und E10, den Identifier der Tankstelle *stid*, und den Zeitstempel *date* übernommen. Zudem wurde unter *id* ein Zähler über alle Änderungen, sowie mit *changed* ein Identifier für die geänderten Kraftstoffsorten, hinzugefügt. Diese Identifier errechnet

sich aus der Summe der entsprechenden Felder der kleinen Tabelle und gibt nicht nur an ob ein Preis geändert wurde, sondern auch, ob dieser auf Null gesetzt (removed), oder von Null wieder erhöht (new) wurde.

2.2 PSQL Datenbank

2.2.1 Funktion

2.2.2 Indexing

2.3 Verwendete Programme und Packages

2.4 Pricing-Tool

Neben den Verbraucher kommen diese Informationen auch den Tankstellenbesitzern zu Gute. Die einheitlich definierte technische Informationsschnittstelle ermöglicht es Preise automatisch ohne Zeitverzögerung anzupassen. Die Homogenität des Produktes, der für den Kunden einzig interessante Unterschied im Bezug auf das Produkt ist der Preis selber, sorgt gleichzeitig dafür, dass Preisanpassungsregeln eine relativ einfache Struktur aufweisen können. Aufgrund dieser Simplizität, sind in den letzten Jahren einige Pricing-Tools entstanden die das Pricingverhalten zumindest teilweise automatisieren. Während mittelständische Unternehmen auf externe Software Lösungen zurückgreifen, so wie zum Beispiel die Q1 und Westfalen AG auf Angebote der Firmen Temiz4u¹ und Weat², ist anzunehmen, dass zumindest die größten fünf Mineralölkonzerne Deutschlands, Aral, Shell, Jet, Esso und Total, intern entwickelte Programme verwenden dürften. Im folgenden werden am Beispiel eines Tools die Einstellungsmöglichkeiten für den Nutzer erklärt, die diesem bei der Definition seiner Regeln zur Verfügung stehen. Diese Optionen stellen gleichsam die Parameter dar, die im Zuge dieser Arbeit aus dem Datensatz extrahiert werden sollen.

2.4.1 Funktionsweise

Im folgenden werden die für diese Arbeit relevanten Funktionen eines solchen Tools beschrieben. Es wird für diese Arbeit angenommen das andere Tools inhaltlich ähnlich funktionieren. Das hier vorgestellte Tool erhält über einen oder mehrere Verbraucherinformationsdienst die aktuellen Preisänderungen aller Tankstellen. Beim eintreffen einer Änderung wird überprüft, ob diese Änderung eine von dem Nutzer hinterlegte Regel verletzt. Wird eine Regel verletzt wird darauf in der wiederum vom Nutzer gewählten Art und Weise reagiert. Die Nutzereinstellungen lassen sich in die drei Teilaspekte, Konkurrenten bestimmen, Regel festlegen und Automatisierungsgrad wählen gliedern.

¹Quelle: <https://www.temiz4u.net/web/guest/solutions>

²Quelle: <http://www.weat.de/produkte/pricing/>

2.4.2 Konkurrenz bestimmen







P4u aktiv	Modus	Typ	Modus erw. ungenannt	Logo	Wettbewerber
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		Aral Osnabrück (49078), Kurt-Schumacher
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		JET Osnabrück (49090), Hansastr. 43a
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		JET Osnabrück (49082), Iburger Str. 193
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		Kaufland Osnabrück (49078), Kurt-Schuma
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		SCORE Osnabrück (49078), Kurt-Schuma
<input checked="" type="checkbox"/>	Erweitert	Mitbewerber	Preisvorschlag		Shell Osnabrück (49078), Lengericher Lan

Abbildung 3: Pricing Tool Konkurrenz Bestimmung

Hier wird eine Liste von Tankstellen zusammengestellt, die als Konkurrenten erachtet werden. Diese Liste dient als erster Filter für die einkommenden Änderungen. Es werden im weiteren nur Änderungen überprüft, deren Tankstelle in dieser Liste vorzufinde ist.

2.4.3 Regeln festlegen


Aral Osnabrück (49078), Kurt-Schumacher-Damm 12										
PM	ID	Produkt	PM	Wettbewerberprodukt	Währung	Preis-Abstand	Wochentag	Zeit von	Zeit bis	RZ Senkung
	21	Super (ROZ95)		Super (ROZ95)	EUR	0,000	Täglich	00:00:00	23:59:59	10
	10	Diesel		Diesel	EUR	0,000	Täglich	00:00:00	23:59:59	10

Abbildung 4: Pricing Tool Regel Festlegung

Für jeden Konkurrenten kann ein eigenes Regelwerk angelegt werden. Jede Regel legt zunächst einen untersten Preisabstand für eine bestimmte Spritsorte fest, welcher zu dem jeweiligen Konkurrenten bewahrt werden soll. Positive Werte besagen, dass die Tankstelle selber maximal um den genannten Betrag teurer sein will als der entsprechende Konkurrent. Negative Werte besagen dementsprechend, dass die Tanstellen selber mindestens um diesen Betrag günstiger sein möchte als ihr Konkurrent. Bei dem Beispiel im Bild wären also gleiche und günstigere eigene Preise erlaubt. Zusätzlich können für den gewählten Abstand die Wochentage sowie die jeweilige Zeit gewählt werden an denen der Abstand einzuhalten ist. Wird der Abstand außerhalb dieses Zeitintervalls unterschritten, hat dies keine Konsequenzen. Im letzten Feld kann eine Reaktionszeit in Minuten angegeben werden. Diese besagt wieviele Minuten nach dem Eingang einer Regelwidrigen Änderung die gewählte Reaktion erfolgen soll.

2.4.4 Automatisierung wählen

Wettbewerber: Aral Osnabrück (49078), Kurt-Schumac ▼

Einstufung: Mitbewerber Erw. Quellenverwertung aktiv: Ja

AI	Preismeldene Quelle	Maßnahme	Wochentag	Zeit von	Zeit bis	Aktic
<input checked="" type="checkbox"/>	Q1 ZETAS ISO Bitmap Meldungen: 2 ▼	SenkungAutomatik ▼	Freitag ▼	16:00:00	17:59:59	▼
<input type="checkbox"/>	Q1 ZETAS ISO Bitmap Meldungen: 2 ▼	Verwerfen ▼	Taeglich ▼	00:00:01	23:59:58	▼
<input checked="" type="checkbox"/>	Q1 ZETAS ISO Bitmap Meldungen: 2 ▼	SenkungAutomatik ▼	WERKTAG ▼	00:00:00	07:00:00	▼
<input checked="" type="checkbox"/>	Q1 ZETAS ISO Bitmap Meldungen: 2 ▼	SenkungAutomatik ▼	WERKTAG ▼	18:00:00	23:59:59	▼
<input checked="" type="checkbox"/>	Q1 ZETAS ISO Bitmap Meldungen: 2 ▼	SenkungAutomatik ▼	WOCHENE ▼	00:00:00	23:59:59	▼
<input checked="" type="checkbox"/>	pricing4u.net Portal (Preismeldungen): 3 ▼	SenkungAutomatik ▼	Freitag ▼	16:00:00	17:59:59	▼
<input type="checkbox"/>	pricing4u.net Portal (Preismeldungen): 3 ▼	Verwerfen ▼	Taeglich ▼	00:00:01	23:59:58	▼
<input checked="" type="checkbox"/>	pricing4u.net Portal (Preismeldungen): 3 ▼	SenkungAutomatik ▼	WERKTAG ▼	00:00:00	07:00:00	▼
<input checked="" type="checkbox"/>	pricing4u.net Portal (Preismeldungen): 3 ▼	SenkungAutomatik ▼	WERKTAG ▼	18:00:00	23:59:59	▼
<input checked="" type="checkbox"/>	pricing4u.net Portal (Preismeldungen): 3 ▼	SenkungAutomatik ▼	WOCHENE ▼	00:00:00	23:59:59	▼
<input checked="" type="checkbox"/>	pricing4u.net Marktdaten D: 777777 ▼	SenkungAutomatik ▼	Freitag ▼	16:00:00	17:59:59	▼
<input type="checkbox"/>	pricing4u.net Marktdaten D: 777777 ▼	Verwerfen ▼	Taeglich ▼	00:00:01	23:59:58	▼
<input checked="" type="checkbox"/>	pricing4u.net Marktdaten D: 777777 ▼	SenkungAutomatik ▼	WERKTAG ▼	04:00:00	07:00:00	▼
<input checked="" type="checkbox"/>	pricing4u.net Marktdaten D: 777777 ▼	SenkungAutomatik ▼	WERKTAG ▼	18:00:00	19:00:00	▼
<input checked="" type="checkbox"/>	pricing4u.net Marktdaten D: 777777 ▼	SenkungAutomatik ▼	WOCHENE ▼	04:00:00	19:00:00	▼

Abbildung 5: Prancing Tool Automatik Einstellung

Das Tool bietet neben einer *default* Variante die beiden weiteren Möglichkeiten *SenkungAutomatik* und *Verwerfen* an, welche auf unterschiedliche Weise auf Regelverletzungen reagieren. Bei der *default* Variante wird bei einer Regelverletzung nach der festgelegten Reaktionszeit eine eigene Preisänderung vorgeschlagen, die den niedrigsten tolerierten Preisabstand wieder herstellt. Der Vorschlag kann dann vom Nutzer nach eigenem Ermessen durchgeführt, abgeändert oder gänzlich verworfen werden. Bei der automatischen Senkung wird statt dem Vorschlag die entsprechende eigene Preisänderung nach der festgelegten Reaktionszeit automatisch durchgeführt. Die letzte Möglichkeit verwirft die entsprechende Änderung ohne den Nutzer darüber zu informieren. Wie auch schon bei den Regeln kann hier für jeden Wochentag für jedes beliebige Zeitintervall eingestellt werden, wie auf einen Verstoß reagiert werden soll.

3 Anforderungsanalyse

Thema dieses Abschnittes ist es, das Ziel der Arbeit genau zu spezifizieren und zu modularisieren. Das heißt das Ziel soll bereits in logische Untereinheiten unterteilt werden. Dafür müssen bereits absehbare Probleme erkannt und notwendige Zwischenergebnisse definiert werden.

3.1 Zielsetzung

Ziel ist es in den Pricingdaten Verhaltensmuster zu erkennen. Es soll in dieser Arbeit ausschließlich um das Erkennen von reaktivem Verhalten gehen. Das heißt, es geht darum zu erkennen nach welchen anderen Tankstellen sich eine Tankstelle in welchem Maße richtet. Dabei wird sich stark an dem Konzept des vorgestellten Tools orientiert. Es ist anzunehmen, dass Tankstellen, die nicht speziell dieses Tool verwenden, entweder ein sehr ähnlich gestaltetes Tool betreiben oder aber manuell nach ähnlichen Prinzipien operieren. Das liegt an den Grundlagen des Wettbewerbes. Für eine Tankstelle geht es darum, einen möglichst hohen Umsatz an Kraftstofflitern zu einem möglichst hohen Preis zu erzielen. Die für den Kunden interessanten Faktoren sind hierbei in erster Linie der Preis und die Nähe, wobei sich die Gewichtung dieser Faktoren zwischen verschiedenen Kunden und zu verschiedenen Zeiten unterscheidet. Das generelle Kundenaufkommen ist abhängig vom Verkehrsfluss, unterscheidet sich also täglich wie stündlich und teilweise auch saisonbedingt. Es sollte also im Interesse einer Tankstelle liegen, zu den konkurrierenden Tankstellen einen bestimmten Preisabstand zu wahren, den es unter Umständen zeitlich zu variieren gilt. Demnach sind die Parameter nach denen in dieser Arbeit gesucht werden soll die jeweilige Konkurrenten einer Tankstelle sowie die gewünschten Preisabstände innerhalb bestimmter Zeitintervalle, was ziemlich genau den Einstellungsoptionen des Tools entspricht.

Ein weiterer Grund sich an den Regeln der Tools zu orientieren besteht darin, dass an tatsächlich in den Tools verwendeten Regeln die Performance des verwendeten Ansatzes überprüft werden kann. Das bedeutet nicht, dass das Ziel dieses ersten Ansatzes bereits sein soll die verfügbaren Regelsätze zu hundert Prozent nachzubilden. Es geht immernoch primär darum, die aus dem tatsächlichen Verhalten ersichtlichen Reaktionsmuster zu erkennen. Das dies ein Unterschied ist wird im weiteren Verlauf deutlich werden. Die Regeln sollen zunächst nur dabei helfen auf ansatzweise valide Ergebnisse zu testen.

3.2 Generelle Voranalyse

Im Bereich der Mustererkennung gibt es ein paar kritische Faktoren, welche besonders ausschlaggebend sind für das prinzipielle Vorgehen. Zunächst muss die Menge der Daten

evaluiert werden. Während bei großen Datenmengen die Geschwindigkeit und der verbrauchte Speicher zu kritischen Faktoren werden, sinkt bei geringen Datenmengen schnell die Aussagekraft. Zudem haben Anzahl, Abgrenzung und Konsistenz der Muster einen großen Einfluss auf die benötigte Datenmenge. Es muss auch geklärt werden welche Daten in welcher Form benötigt werden und wie mit fehlenden und fehlerhaften Werten umgegangen werden soll. Zuletzt müssen anhand der Beschaffenheit des Problems, der Verfügbarkeit von Trainingsdaten und der Auswertung der genannten Faktoren mögliche Lösungsansätze ermittelt werden.

3.2.1 Datenmenge

Der Datensatz besteht aus knapp 15000 Tankstellen und insgesamt gut 80 Millionen Preismeldungen, erfordert also aufgrund seiner Größe eine Vorstrukturierung. Eine erste Selektion der Daten ist allein durch die Struktur der gesuchten Muster bereits gegeben. Es werden preisliche Korrelationen zwischen jeweils zwei Tankstellen vermutet, was die Menge durchschnittlich auf circa 5000 Preisänderungen pro Tankstelle reduziert. Hinzu kommt, dass konkurrierende Tankstellen nicht nur eine Regel haben müssen nach der sie aufeinander reagieren. In Anlehnung an die Funktionen des Pricing Tools wäre es möglich, dass verschiedene Regeln nach einiger Zeit wechseln, für bestimmte Tage definiert sind oder sogar nur für bestimmte Uhrzeiten.

Einmal angenommen eine Tankstelle würde für ein halbes Jahr eine Regel für eine Stunde für einen bestimmten Wochentag festlegen. Dann würden durchschnittlich lediglich knapp acht Änderungen in den relevanten Zeitraum fallen. Weiterhin seien alle diese Änderungen reaktive Preissenkungen und die Tankstelle hätte vier gleichwertige Konkurrenten. Dann wären es noch zwei Preisänderungen, die dieser Regel entsprechen würden. Abgesehen davon, dass die Tankstelle auch noch die Möglichkeit hat die eigenen Regeln zu ignorieren, wären das viel zu wenig Daten um eine statistisch signifikante Aussage treffen zu können. Hier ist auch noch nicht berücksichtigt, dass auch genau diese zwei Änderungen als Reaktionen auf diesen Konkurrenten erkannt werden müssen.

Die hohe Anzahl der durch die Flexibilität in den Zeitintervallen theoretisch möglichen Regeln macht es also notwendig ein paar Einschränkungen zugunsten signifikanterer Ergebnisse vorzunehmen.

3.2.2 Strukturierung des Datensatzes

Die Datenanalyse soll in Python durchgeführt werden. Die Daten müssen also zunächst von der Datenbank importiert werden. Aufgrund der Menge der Daten, sollten nur diejenigen importiert werden, die für die Analyse einer Tankstelle benötigt werden, um den Arbeitsspeicher nicht zu überlasten. Die Daten sollten so in Python abgespeichert werden,

dass besonders einfach und schnell auf die benötigten Datensätze zugegriffen werden kann. Dazu muss herausgearbeitet werden, nach welchen Kriterien besonders häufig auf Daten zugegriffen wird. Es muss auch untersucht werden, ob zusätzliche Daten notwendig oder Datenfelder unnötig sind. Auch die jeweiligen Datentypen der einzelnen Felder müssen eventuell geändert werden.

Da die Daten zu einer Preisänderung von den Tankstellen selber angegeben werden, gibt es ein paar strukturelle Unterschiede in den Angaben. Zunächst ist es möglich pro Kraftstoffsorte eine eigene Preisänderung vorzunehmen, oder aber die Änderungen in Gruppen oder alle zusammen in einem Pricing durchzuführen. Dann gibt es Tankstellen, die vor jeder Änderung den Preis zunächst erst einmal auf Null setzen, bevor sie ihn wieder auf den eigentlich gewünschten Stand anheben. Auch senken einige Tankstellen den Preis bei Tagesende auf Null herab. In einigen Fällen werden Preise für die selbe Kraftstoffsorte in kürzester Zeit mehrmals geändert. Alle diese Vorgehensweisen haben potentiell verschiedene Auswirkungen auf eine automatische Analyse. Es muss einerseits überprüft werden, ob Differenzen unnötig sind und im Vorhinein angeglichen werden können oder aber wichtige Informationen enthalten und deshalb bewahrt und verschiedenartig behandelt werden müssen. Andererseits muss auch untersucht werden inwiefern diese Auswirkungen im späteren Verlauf zu Problemen führen könnten. Letzteres sollte auch im Bezug auf fehlerhafte, fehlende oder wenig sinnvolle Daten geschehen.

3.2.3 Algorithmen und andere Ansätze

Generell geht es bei der Aufgabe darum zu entscheiden, ob eine bestimmte Tankstelle auf eine andere reagiert. Demnach ist es unter anderem auch ein Klassifizierungsproblem. Die konkreten Regeln selbst können nicht eindeutig kategorisiert und klassifiziert werden, da sie in ihrer Form zu variabel sind. Allerdings resultieren die Regeln in einzelnen Reaktionen der Tankstellen aufeinander. Man könnte versuchen zu klassifizieren ob eine einzelne Preisänderung die Reaktion auf eine bestimmte andere Preisänderung ist. Im anschließend resultierenden Set an Reaktionen, könnte dann nach den spezifischen Regeln gesucht werden.

Eine Möglichkeit Reaktionen zu klassifizieren wäre es Methoden aus dem Bereich des maschinellen Lernens zu benutzen. Diese Methoden benötigen zum Erstellen eines Modells, welches über die einzelnen Daten entscheidet, klassifizierte Trainingsdaten. Als Trainingsdaten stehen nur wenige exemplarische Regeln eines einzigen Tankstellenunternehmens zur Verfügung. Man könnte mittels dieser Regeln einige Reaktionen aus dem Datensatz extrahieren. Dies wäre jedoch kein repräsentativer Auszug aus dem Datensatz, da einerseits zu wenige Regelsätze zur Verfügung stehen und andererseits nur das Verhalten eines

Anbieters enthalten ist. Das Model würde also selbst wenn die Menge der Trainingsdaten ausreichend wäre nur die Charakteristiken im erhalten einer Tankstelle lernen und wäre damit nicht auf die restlichen Daten übertragbar.

Als Alternative dazu könnte man die Reaktionen als kausale Verkettungen betrachten und die entsprechenden Zeitreihendaten mittels statistischer Methoden auf Korrelationen untersuchen. Zur Analyse von Kausalität in Zeitreihendaten werden oftmals verschiedene Variationen des Granger Causality Tests verwendet. Bei diesem Test wird generell geprüft ob zwei Zeitreihendaten in kausaler Relation stehen. Zeitreihe A wäre kausal abhängig von Zeitreihe B, wenn die Vorhersage des jeweils nächsten Wertes in Zeitreihe A signifikant verbessert werden würde, wenn man als Parameter neben den vorherigen Werten von A auch die vorherigen Werte aus der Zeitreihe B hinzunimmt. Für einen solchen Test müssten zunächst Zeitreihendaten generiert werden. Da Reaktionen innerhalb von kürzester Zeit erfolgen können müssten geringe Zeitintervalle gewählt werden. Das würde zu Zeitreihen mit fast ausschließlich gleichen Werten führen, weil der Preis oft über mehrere Stunden gleich bleibt. Ein Vorhersagemodel, was immer den gleichen Preis vorhersagt wäre also nicht mehr signifikant zu übertreffen. Man könnte versuchen dieses Problem zu umgehen und die Zeitintervalle in denen keine Änderung stattfindet ausschneiden. Allerdings wollen Tankstellen vermutlich lediglich einen bestimmten Preisabstand waren und reagieren demnach nicht konsistent auf absolute Preise, sondern nur auf den Anteil der eine bestimmte Preisdifferenz aufweist. Zudem könnten unterschiedliche Abstände sich immer wieder ablösen, sodass zunächst die Daten verschiedenartig in unterschiedliche Zeitfenster unterteilt werden müsste.

Bei dieser Vielzahl an Faktoren nach denen die Daten vorselektiert werden müssten, machen die meisten generellen Algorithmen wenig Sinn. Eine eigens auf diese Problemstellung zugeschnittene logische Struktur zu entwickeln scheint am zielführensten. Die Regeln sind inhaltlich über einen Preisabstand definiert. Es sind also nicht absolute Preise, sondern die Differenzen zu den Konkurrenten die wirklich entscheidend sind. Es sollte also eine detailliert Analyse über diese Differenzen durchgeführt werden. Auch ist nicht jeder Differenzwert gleich interessant. Besonders wichtig sind im Konkurrentenverhältnis die vergleichsweise hohen Preisabstände. Dabei ist hier nicht der absolute Betrag gemeint. Die Differenz wird aus dem Blickwinkel der reagierenden Tankstelle errechnet, wobei negative Werte bedeuten, dass diese Tankstelle einen niedrigeren Preis aufweist als der jeweilig betrachtete Konkurrent. Hohe Differenzen bedeuten für die reagierenden Tankstellen als einen verhältnismäßig ungünstigen Preisabstand, wenn mit ungünstig hier die Attraktivität für Kunden gemeint ist. Der in der Regel festgelegte Preisabstand bildet also einen maximalen Schwellenwert. Bei Überschreitung, wird der eigene Preis auch angepasst um den maximalen Schwellenwert wieder herzustellen. Die maximal vorkommenden Differen-

zen sind also, sofern keine Ausreißer, der Schwellenwert der Regel. Dabei ist es genauso wichtig, die Preisänderungen der Konkurrenten auf welche nicht reagiert wurde zu untersuchen, wie auf die tatsächlichen Reaktionen. Auch hier bilden die maximalen Werte den Schwellenwert der Regel. Dabei muss zusätzlich berücksichtigt werden, dass für unterschiedliche Zeitintervalle unterschiedlich hohe Schwellenwerte hinterlegt seien könnten. Eine detaillierte Analyse der Differenzen zweier Tankstellen ist vor dem Hintergrund dieser Vorüberlegungen vom Ansatz her am vielversprechendsten.

3.3 Modularisierung

Aufgrund der großen Datenmenge sollten die Tankstellen einzeln analysiert werden können, anstatt alle in einem Durchlauf abzuarbeiten. Auch wenn es eventuell möglich wäre die Konkurrenten einer Tankstelle einzeln zu analysieren, ist es unter Umständen hilfreich alle potentiellen Konkurrenten gleichzeitig zu untersuchen. Da nur jeweils eine Preisänderung pro eigenem Pricing als Auslöser in Frage kommt, könnten so verschiedene potentielle Auslöser verglichen werden. Aus diesem Grund wird die Analyse einer Tankstelle auch nicht direkt zweiseitig durchgeführt. Damit wäre gemeint, für eine Tankstelle sowohl die Reaktionen auslösenden Tankstellen, als auch die auf diese Tankstelle reagierenden Tankstellen zu ermitteln. In Verbindung mit dem Ansatz alle potentiell auslösenden Tankstellen gleichzeitig zu untersuchen würde eine zweiseitige Betrachtung zu größeren Verkettungen führen, da bei der Analyse der reagierenden Tankstellen auch wieder deren komplettes Set an möglichen Auslösern verglichen werden müsste. Deshalb werden nur die Reaktionen auslösenden Konkurrenten eines Standortes bestimmt. Zudem kann ein Zeitraum angegeben werden, für den die Analyse einer Tankstelle durchgeführt werden soll.

Um nicht alle anderen 15000 Tankstellen anhand ihres tatsächlichen Verhaltens überprüfen zu müssen, sollte eine erste Vorauswahl getroffen werden. Alle wirklichen Konkurrenten auszuwählen hat dabei höhere Priorität als die Auswahl gering zu halten. Anschließend müssen für jede eigene Preisänderung die möglichen Auslöser der potentiellen Konkurrenten ermittelt werden. Zudem werden bei jedem potentiellen Konkurrenten die Preisänderungen ermittelt, auf die nicht reagiert wurde, sowie der jeweilige eigene Preisstand zu diesem Zeitpunkt. Für die Summe aller Preispaarungen pro Tankstellenpaar wird dann die Verteilung der Preisdifferenzen gebildet. Von den Maximalwerten abwärts wird dann auf mögliche Regeln überprüft.

4 Umsetzung

4.1 Datenaufbereitung

Mit Hilfe des Python Modules *psycpg2* werden die benötigten Daten aus der Datenbank nach Python importiert.

4.2 Tankstellendaten

Der komplette Datensatz an Stammdaten zu den einzelnen Tankstellen wird als Instanzen der Klasse *station* in einem *dictionary* abgelegt, sodass jede Tankstelle über ihre *id* abgerufen werden kann. Jedes einzelne Feld wird entweder übernommen oder, falls nicht verfügbar, mit einem Defaultwert gekennzeichnet. Der Großteil der benötigten Funktionalitäten wird über diese Klasse *station* bereitgestellt.

4.2.1 Preisdaten

Die für die Analyse benötigten Preisdaten einer Tankstelle werden in einer Matrix chronologisch gespeichert. Bis auf die Tankstellen-Id werden alle Felder übernommen. Der Zeitstempel wird in Zahlenwerte für das Datum und die Uhrzeit umgerechnet. Das Datum besteht in dem Abstand zum Datum der ersten Änderung in Tagen und die Uhrzeit in den an dem Tag vergangenen Sekunden. Zu dem Datum werden zusätzlich der Wochentag und der Monat bestimmt. Zu den jeweiligen aktuellen Preisen werden auch die Änderungsbeträge für jede Kraftstoffsorte ermittelt. Jede Form der Preisänderung wird bis auf einen Ausnahmefall so erhalten, wie sie ist. Die Struktur ist von Bedeutung, wenn es darum geht inhaltlich zu bestimmen, welche Änderungen potentiell ein Auslöser sein könnten. Die Ausnahme besteht darin, wenn Preise auf Null herabgesenkt werden. Diese Fälle würden extreme Differenzwerte erzeugen, welche in der Ermittlung der Schwellenwerte für die Regeln Fehler erzeugen könnten. Außerdem stellen sie inhaltlich keine wirkliche Preisänderung dar. In diesen Fällen wird für die entsprechende Kraftstoffsorte der bis dahin geltende Preis genommen. Würde durch diese Abwandlung eine Pricing entstehen, das keine Änderung beinhaltet, so wird der komplette Eintrag entfernt.

4.3 Potentielle Konkurrenten ermitteln

Wie in der Vorüberlegung angedeutet, soll hier zunächst eine grobe Auswahl der überhaupt als Konkurrenten in Frage kommenden anderen Tankstellen getroffen werden. Das Hauptkriterium hierfür soll der Abstand zur untersuchten Tankstelle sein, welcher über die Geokoordinaten bestimmt werden kann. Dafür wird der Einfachheit halber der direkte Abstand über die Erdoberfläche errechnet. Es ist allerdings nicht ausreichend einfach einen

maximalen Abstandswert festzulegen und alle im Umkreis liegenden Tankstellen als potentielle Konkurrenten zu betrachten. Die Anzahl der umliegenden Tankstellen schwankt stark mit dem Besiedelungsgrad der unmittelbaren Umgebung und somit auch die maximalen Abstände von tatsächlichen Konkurrenten. So kann es in der Stadt durchaus vorkommen, dass der nächste Konkurrent auf der anderen Straßenseite zu finden ist und sich noch 15 oder mehr weitere Tankstellen im Umkreis von wenigen Kilometern befinden, wohingegen es beispielsweise auf der Autobahn vorkommen kann, dass die nächste Tankstelle mehr als zehn Kilometer entfernt liegt.

In der aktuellen Version wird deshalb eine Mischung aus einer Variablen Abstandsobergrenze, einer Mindestanzahl und einer maximalen Anzahl an potentiellen Konkurrenten verwendet. Solange im Umkreis der Obergrenze nicht genügend Tankstellen gefunden wurden, wird die Obergrenze von anfänglich fünf Kilometern immer wieder verdoppelt. Sollte die maximale Anzahl überschritten werden, so werden von den am weitesten entfernten die Überzähligen entfernt. Die so ermittelten potentiellen Konkurrenten werden als Liste ihrer *ids* nach dem Abstand geordnet in der gerade analysierten Station abgespeichert. Grund für dieses Vorgehen anstatt lediglich einer hohen Mindestzahl ist das nicht unnötig viele Tankstellen ausgewählt werden sollen. Bei jeder Tankstelle besteht ein relativ hohes Risiko Korrelationen zu finden, die über Kettenreaktionen und Interaktionen mit dritten Tankstellen hervorgerufen werden. Weitere Möglichkeiten werden im Fazit kurz angesprochen.

4.4 Reaktionen erkennen

Grundsätzlich wird hier durch die Preisänderungen der untersuchten Tankstelle und der potentiellen Konkurrenten durchiteriert und nach den im folgenden beschriebenen Kriterien Reaktion und deren Auslöser extrahiert. Dabei werden Preiserhöhungen, sowie Preissenkungen kurz nach einer Erhöhung, nicht miteingeschlossen. Anstatt Kopien der Änderungen selber werden die Indexe der Paarungen in ihren Matrizen in Listen gespeichert, einerseits um Arbeitsspeicher zu sparen und andererseits um den historischen Kontext schnell herleiten zu können. Dabei wird je eine Liste für die potentiellen Konkurrenten erstellt um schnell und einfach die Statistiken für die Regeln ableiten zu können, sowie eine chronologisch sortiert nach den Reaktionen mit Unterlisten für alle potentiellen Auslöser um diese später Untereinander vergleichen zu können.

Es geht an dieser Stelle noch nicht darum ausschließlich die genauen Auslöser der Preissenkungen einer Tankstelle zu finden. Dazu gibt es nicht ausreichend eindeutige Informationen. Es geht vielmehr darum alle potentiellen Auslöser zu sammeln um darüber anschließen mit statistischen Mitteln die wahrscheinlichsten Konkurrenten und die mit ihnen verbundenen Regeln ermitteln zu können. Dabei ist es unumgänglich Änderungen

als Auslöser zu interpretieren, die prinzipiell in Frage kämen, jedoch tatsächlich nicht wirksam wurden, weil die festgelegte Preisschwelle durch sie nicht unterschritten wurde, da diese Schwelle noch nicht bekannt. Diese Fälle sind aber auch nicht besonders kritisch, da die maximalen Differenzen zu einem potentiellen Konkurrenten zuerst untersucht werden.

Probleme können entstehen, wenn auf einem engen Zeitraum mehrere Senkungen vollzogen werden und falsche Paarungen von Reaktionen und Auslösern gemacht werden, weil so potentiell systematisch höhere Differenzwerte als die Schwellenwerte entstehen könnten, die dann statistisch signifikant werden könnten. Deswegen wird bei solchen Fällen eine extra Fallunterscheidung durchgeführt. Ausreißer kann es durch bewusste Regelverletzungen oder technische Ausfälle natürlich trotzdem geben. Diese sollten jedoch so sporadisch sein, dass man sie als eben solche Ausreißer erkennen können sollte. Zunächst werden Reaktionen jedoch nach einer zeitlichen und einer semantischen Komponenten ausgewählt.

4.4.1 Zeitliche Komponente

Auslöser und Reaktion sollten zeitlich nahe beieinander liegen. Das heißt es muss ein maximales Zeitfenster vor einer Preisänderung festgelegt werden, in welchem ein Pricing von einem potentiellen Konkurrenten als Auslöser in Frage kommt. In den letzten Monaten gab es häufig sechs bis acht Preisänderungen verteilt über ein Zeitintervall von weniger als 16 Stunden. Bei Reaktionen mit mehr als zwei Stunden Verzögerung wäre es also nicht unwahrscheinlich, dass die auslösende Tankstelle bereits eine erneute Senkung durchgeführt hat, bevor die Reaktion eintritt. Selbst bei einer Stunde würde bereits insgesamt mehr Zeit bei einem eigentlich ungewünschten Preisabstand verbracht werden als bei dem eigentlich angestrebten. Stark verzögerte Reaktionen sollten somit aus ökonomischer Sicht vermieden werden, um beim Kunden den Eindruck zu vermeiden, dass die Tankstelle verhältnismäßig teurer ist.

Trotzdem kann bei automatischen Pricing Tools eine Reaktionszeit eingestellt werden. Die in den Testdaten verwendeten Zeiten sind nicht größer als zehn Minuten. Andere Tankstellen könnten aber durchaus auch höhere Reaktionszeiten eingestellt haben. Tankstellen, die keine Pricing Tools benutzen, haben höchstwahrscheinlich generell eine höhere Reaktionszeit. Zeitenfenster unter 20 Minuten zu verwenden, würde also mit ziemlicher Sicherheit einige Reaktionen nicht berücksichtigen. Um dem vorzubeugen wurde ein Zeitfenster von 45 Minuten gewählt. Ein zu großes Zeitfenster würde dazu führen mehr korrelierende, aber nicht kausal verantwortlich, Preisänderungen in den Pool der potentiellen Auslöser aufzunehmen.

4.4.2 Semantische Komponente

Inhaltlich sollte der Höhenbetrag der Änderung in allen geänderten Kraftstoffsorten bei der Reaktion kleiner oder gleich der Änderung des Auslösers sein. Das liegt daran, dass der Preisabstand vor einem Auslöser maximal dem Schwellenwert entspricht. Ein Auslöser kann den Schwellenwert also maximal um die Höhe seiner Änderung unterschreiten. Eine Reaktion die den Schwellenwert wieder herstellt darf also nicht höher ausfallen als der Auslöser selber. Hier wird auch deutlich warum die verschiedenen Strukturen des Pricings beibehalten werden sollten und nicht etwa ein Datensatz für jede Kraftstoffsorte erstellt wurde. Entspricht nur eine Kraftstoffsorte einer Preisänderung nicht notwendigen Kriterien, so kommt die ganze Änderung als regelkonforme Reaktion nicht mehr in Frage. Bei einzelner Betrachtung der Kraftstoffsorten wäre diese Verschärfung des Ausschlusskriterium nicht möglich gewesen.

4.4.3 Fallunterscheidung

Es gibt vier verschiedene Fälle, die durch die verschiedenartigen Strukturen im Pricing Verhalten aber noch verkompliziert werden. Generell soll es möglich sein, dass ein Auslöser mehrere Reaktionen hervorruft und zwar maximal eine pro Kraftstoffsorte. Umgekehrt gilt das Gleiche, also das mehrere Auslöser eine Reaktion hervorrufen. Das liegt daran, dass es denkbar wäre dass sowohl die verantwortlichen Personen als auch automatische System Pricings mit mehreren Kraftstoffsorten in einer Änderung in ihre Einzelteile aufspalten oder auch Einzelne zu einem Ganzen zusammenfügen. Es soll aber ausgeschlossen sein, dass pro Auslöser-Reaktion-Paarung ein Kraftstoff bei einer Partei mehrmals geändert wird. Bei manuellen Änderungen ist es zwar möglich, dass bei einer Anpassung Fehler unterlaufen und deswegen eine weitere hinzugefügt werden muss, doch sollten diese Fälle so vereinzelt sein, dass es das Ergebnis nicht beeinflussen sollte. Auslöser-Reaktion-Paarungen können also maximal drei Änderungen pro Seite enthalten. Zudem können in einem Zeitfenster verschieden viele unabhängige Preisänderungen auf beiden Seiten stattfinden. Deshalb werden pro Seite die beiden Kategorien eine Preisänderung oder mehrere unterschieden, sodass sich aus den verschiedenen Kombinationen insgesamt vier verschiedene Fälle unterscheiden lassen.

ein Auslöser - eine Reaktion Die am häufigsten vorkommende Variante. Hier muss nur die semantische Komponente überprüft werden.

ein Auslöser - mehrere Reaktionen In diesem Fall befindet sich nur eine einzelne potentiell auslösende Preisänderung aber mehrere Preisänderungen der untersuchten Tankstelle in einem Zeitintervall. Die letztlich gewählte Reaktion sollte auf jeden

Fall das semantische Kriterium erfüllen. Bei den späteren Reaktionen muss für die Überprüfung dieses Kriteriums die Summe aller vorherigen Änderungen in diesem Zeitintervall auf die jeweilige Reaktion selber draufaddiert werden. Denn falls eine der späteren Änderungen die tatsächliche Reaktion war, dann war sie trotz allen vorherigen, ohnehin schon durchgeführten, Änderungen anscheinend immernoch erforderlich. Es wird hierbei davon ausgegangen, dass auch die automatischen Tools nach dem Abwarten der Reaktionszeit vor der Durchführung der eigenen Änderung noch einmal den zu ändernden Betrag überprüfen. Würden sie dies nicht tun, könnte das dazu führen, dass eine Tankstelle systematisch einen geringeren Preis als nötig fährt und somit Gewinne einbüßt. Deshalb werden die potentiellen Reaktionen, angefangen bei der Ersten, aufaddiert, bis das semantische Kriterium nicht mehr erfüllt wird. Die letzte Änderung die das Kriterium noch erfüllt wird als tatsächliche Reaktion eingestuft. Wenn diese Änderung alleine nicht alle Kraftstoffsorten die beim Auslöser geändert wurden abdeckt, können vorherige noch hinzugenommen werden, sofern sie nicht zu einer doppelten Änderungen einer Sorte führen würden.

Es ist nicht möglich die früheren Reaktionen inhaltlich komplett auszuschließen. Die Früheren zu wählen würde aber einen höheren Schwellenwert unterstützen. Die späteren Reaktionen sind, weil sie vorherigen in dieser Hinsicht implizieren summiert größerer Anpassungen, stellen also letztlich wieder eine niedrigere Differenz her. Es ist also sicherer die Reaktionen zu nehmen die bei gleichzeitigem Erfüllen des semantischen Kriteriums am spätesten sind, um nicht das Risiko einzugehen unsichere größere Schwellenwerte zu begünstigen. Außerdem ist dieses Szenario extrem unwahrscheinlich. Der Auslöser, der mehrere Reaktionen erklären kann, also an sich schon hohe absolute Senkungen umfasst, mindestens zwei Cent, löst nur die frühere Reaktion aus. Es muss also in kürzester Zeit noch einen weiteren Preisverfall geben, der den bisherigen noch einmal unterschreitet um die weiteren Anpassungen zu rechtfertigen. Der Preis würde also in kürzester Zeit um mindestens 3 Cent verfallen. Es sollte im Interesse keiner Tankstelle sein hohe Preisverfälle zu erzeugen. **PROBLEM:** Angenommen es kommt zunächst eine Reaktion die gegen das semantische Kriterium verstößt, weil sie einen Preis ändert der vom Auslöser nicht verändert wurde, so würde gar keine Reaktion gewählt obwohl spätere durchaus die wahren Reaktionen sein könnten.

mehrere Auslöser - eine Reaktion Der umgekehrte Fall, bei welchem mehrere Änderungen eines Konkurrenten einer einzelnen Reaktion vorausgehen, lässt eine eindeutige Entscheidung zu. In diesem Fall ist die letzte semantisch plausible Änderung des Konkurrenten der potentielle Auslöser. Angenommen eine vorherige Änderung hätte die Reaktion ausgelöst, dann hätten alle noch folgenden Änderungen wieder-

um Reaktionen auslösen müssen. Es erfolge jedoch offenbar nur eine. Zudem ist diese Lösung wieder die sichere Variante. Bei den späteren Änderungen sind schon mehr Senkungen erfolgt, der Konkurrent hat also einen niedrigeren Preis. Die Reaktion stellt hier also einen verhältnismäßig niedrigen Schwellenwert wieder her. Sollte also der Fall eintreten, dass doch eine der früheren Änderungen der Auslöser war und es wurde manuell eine weitere Senkung verhindert, so ist dies zum Einen ein von der Tankstelle gewollter Ausreißer und sollte auch so in den Statistiken auftauchen und zum Anderen auch ein Ausreißer nach unten und somit nicht besonders ausschlaggebend, was die Wahl der Schwellenwerte betrifft.

Auch hier muss bei der Auswahl wieder berücksichtigt werden, dass eventuell nur mehrere Änderungen zusammen den Auslöser bilden. Die Auslöser werden deshalb von hinten durchiteriert. Handelt es sich bei dem aktuellen Auslöser um eine Änderung, bei der es keinerlei Überschneidung bei den geänderten Kraftstoffsorten mit der Reaktion gibt, so kann diese Änderung übergangen werden. Das liegt daran, dass diese Änderung auch nicht eine erneute Änderung hätte auslösen müssen. Dies ist jedoch der Fall, sobald es eine solche Überschneidung gibt. Wenn die Änderung mit dieser Überschneidung das semantische Kriterium noch nicht erfüllt können frühere hinzugenommen werden. Wieder dürfen dabei keine doppelten Änderungen einer Kraftstoffsort entstehen.

mehrere Auslöser - mehrere Reaktion Die letzte Variante ist in gewisser Weise eine Kombination der vorherigen Fälle, was sich auch in der logischen Begründung widerspiegelt. Zunächst wird für die Kombination aller Auslöser untersucht, wieviele Reaktionen dadurch erklärt werden können. Dazu werden die Auslöser zu einem einzelnen künstlichen Auslöser zusammengefasst und dann die Reaktionen der Reihe nach addiert, solange das semantische Kriterium erfüllt bleibt. Sofern am Ende Reaktionen überbleiben, werden diese mit der selben Begründung, wie im 2. Fall außer acht gelassen. Können alle Reaktionen erklärt werden, ist dies nicht notwendig.

Die übrigen Reaktionen werden dann einzeln in rückwärtiger Reihenfolge untersucht und gefunden Paarungen werden für die weitere Untersuchung gelöscht, damit sie nicht doppelt vorkommen. Für diese, durch das Löschen ist sie die jeweils letzte der zu betrachtenden Reaktionen, werden alle davorliegenden Auslöser gesammelt und der 3. Fall wird angewendet. Das Entscheidungskriterium zugunsten des letzten möglichen Auslösers scheint hier außer Kraft gesetzt, denn spätestens ab der zweiten Iteration hatte es ja noch weitere spätere Reaktionen gegeben. Allerdings wurde für diese durch die rückwärtige Analyse bereits ein Auslöser gefunden, oder bewiesen

dass es keinen gibt, wodurch dieses Kriterium doch weiterhin gültig ist.

Wurden ein oder mehrere Auslöser für die aktuelle Reaktion gefunden muss nur noch überprüft werden, ob weitere Reaktionen zwischen dem letzten der Auslöser und dieser aktuell letzten Reaktion liegen, die zusätzlich durch die aktuell gewählten Auslöser erklärt werden können. Es muss nicht noch einmal die Summe wie im 2. Fall gebildet werden, da diese Notwendigkeit durch die zu allererst gebildete Summe behoben wurde.

Dabei kann es rein theoretisch wieder vorkommen, dass Auslöser, die zu einer späteren Reaktion passen, mit diesen gepaart und entfernt werden, obwohl sie eigentlich die Auslöser einer anderen früheren Reaktion sind. Für diese Handhabung sprechen wieder die selben Gründe wie im 2. Fall. Es ist aus ökonomischer Sicht höchst unwahrscheinlich und zudem sicherer eine spätere Reaktion als tatsächliche Reaktion in Betracht zu ziehen, weil dabei weniger kritische Fehler begangen werden.

Wurde kein Auslöser gefunden wird nur die gerade aktuell letzte Reaktion gelöscht.

4.5 Regeln spezifizieren

Aus den gesammelten potentiellen Reaktionen sollen nun die in den Tools hinterlegten Regeln oder aber auch lediglich konsistente Konkurrenzmuster abgeleitet werden. Dazu sollen maximal zugelassene Preisabstände sowie die dazugehörigen Zeitintervalle erkannt werden. Letztlich muss dann über die errechneten Statistiken ermittelt werden, welche Tankstellen wirklich Konkurrenten sind.

4.5.1 Maximaler Preisabstand

Die Verteilung der Preisdifferenz zweier Tankstellen ist das Hauptkriterium auf welches sich die Klassifikation von Konkurrenz stützen wird. Wurde eine Regel und damit ein Schwellenwert für einen Konkurrenten hinterlegt, so sollte die Differenz nie lange über diesen Wert steigen. Es gibt zwei Möglichkeiten, wie eine Tankstelle auf die Änderung eines Konkurrenten entsprechend einer Regel reagieren kann. Entweder gar nicht, weil der Schwellenwert nicht überschritten wurde. Oder aber mit einer eigenen Preissenkung weil eben dies der Fall war. Deshalb werden zwei verschiedene Differenzwerte unterschieden. Zum einen der Preisunterschied nach einer Änderung auf die nicht reagiert werden muss. Der Maximalwert beschreibt somit den maximalen Abstand auf den nicht reagiert werden musste, was gleichzeitig dem Schwellenwert entspricht. Zum Anderen der Preisunterschied nach einer potentiellen Reaktion. Für den Fall, dass es eine tatsächliche Reaktion

und nicht nur eine Korrelation ist, dann sollte nach der reagierenden Preisänderung der Schwellenwert wieder hergestellt worden sein. Bei beiden diesen Differenzen bestimmen die maximalen Werte den potentiellen Schwellenwert einer Regel.

Bei der Analyse dieser Werte werden die einzelnen Kraftstoffsorten zum ersten mal komplett getrennt betrachtet. Werden bei einer Reaktion nicht alle Kraftstoffe angepasst so sollten die Gleichgebliebenen eben nicht in die Kategorie Reaktion, sondern in die der nicht notwendigen Reaktion fallen. Für jeden potentiellen Konkurrenten werden deshalb je zwei Histogramme über die Preisunterschiede pro Kraftstoffsorte erstellt. Das Erste mit Hilfe der Liste aller potentiellen Reaktionen auf einen Konkurrenten. Und das Andere mit allen Preisänderungen eines Konkurrenten die nicht in dieser Liste auftauchen und keine Erhöhungen sind. Jede dafür in Frage kommende Änderung wird noch einmal dahingehend überprüft, ob tatsächlich kein Pricing der untersuchten Tankstelle im anschließenden Zeitfenster zu finden ist. Durch das semantische Kriterium sind schließlich zeitlich korrelierende Anpassungen ausgeschlossen wurden, weil sie keine Reaktionen darstellen, wo der Auslöser jedoch auch nicht notwendigerweise einen Preisabstand hergestellt hat, auf den nicht reagiert werden musste. Es könnte durchaus sein, dass eine Reaktion erforderlich gewesen wäre, hätte nicht sowieso ein anderer Konkurrent eine stärkere Reaktion bewirkt. Zudem müssen für die wirklich ignorierten Änderungen des Konkurrenten noch die gerade geltenden Preise der untersuchten Tankstelle ermittelt werden.

Die beiden Histogramme müssen dann für die Bestimmung des Maximalwertes zunächst zusammengerechnet werden. Der dabei entstehende Maximalwert kann jedoch nicht sofort zum Schwellenwert gemacht werden. Es muss berücksichtigt werden, dass es durchaus noch sein kann, dass Tankstellen die Preise ohne Hilfe eines Tools anpassen, oder zumindest die Anpassung selber noch manuell durchführen. Es kann also vorkommen, dass obwohl eine Regel besteht ab und zu dagegen verstoßen wird. Die Maximalwerte müssen also zunächst auf Ausreißer überprüft werden. Beziehungsweise, es muss überprüft werden, ob die Häufigkeit, mit der ein entsprechender Wert Auftritt regelmäßig genug ist um die Vermutung von konkurrierendem Verhalten zu stützen. Diesbezüglich kann es durchaus vorkommen, dass zu einer Tankstelle eine Regel hinterlegt wurde, welcher aber nur sehr selten eine Reaktion erfordert. Es kann zum Beispiel sein, dass der entsprechende Konkurrent selber in erster Linie auf die gerade untersuchte Tankstelle reagiert. Extremfälle dieser Art können nicht berücksichtigt werden, da das tatsächlicher Verhalten diese Regeln nicht rechtfertigt und es somit eher angeraten wäre die Existenz der Regeln zu hinterfragen.

Zusätzlich gilt es noch zu berücksichtigen, dass sollte ein Wert zurückgewiesen werden, alle Differenzwerte die diesen Wert gestützt hätten zu Regelverstößen werden würden. Es gibt hierzu allerdings Tankstellenübergreifenden keine Erfahrungswerte, wie oft gegen Regeln verstoßen wird. Es ist durchaus möglich, dass sich dieser Faktor zwischen

verschiedenen Unternehmen stark unterscheidet. Grundsätzlich sollte aber davon ausgegangen werden können, dass sich größtenteils an Regeln gehalten wird, da diese Regeln einer ökonomischen Optimierung entstammen und Verstöße generell eine Gefahr bergen Kunden zu verlieren.

Diese beiden Faktoren sprechen dafür schon bei einer recht geringen Anzahl eine Regelmäßigkeit zu vermuten. Als Grenzwertfunktion wurde dafür die Funktion:

$$\frac{2}{\pi} * \arctan(2 * \frac{x^2}{y}) \text{ mit } x=\text{Häufigkeit}, y=\text{Grundgesamtheit}$$

Der Grenzwert sollte relativ zu der Menge der in diesem Zeitintervall liegenden Preisänderungen gewählt werden. Um die Häufigkeit des Wertes stärker zu gewichten, wird diese quadriert, damit bei moderat hohen Häufigkeiten eine unverhältnismäßig höhere absolute Anzahl an Preisänderungen nötig wäre um diesen Wert zurückzuweisen. Angenommen es würde einfach eine Prozhürde von 10% gewählt so würde bei einer passenden Änderung aus zehn bereits ein Regel vermutet wohingegen 9 aus 100 zurückgewiesen werden würden.

4.5.2 Zeitintervall

Eine andere Herangehensweise könnte hier helfen. Man könnte alle Preisänderungen eines Konkurrenten zunächst unterteilen in Änderungen, auf die eine eigene Reaktion stattgefunden hat und jene auf die nicht reagiert wurde. In diesen beiden Gruppen könnte man versuchen die Daten bezüglich verschiedener Parameter immer weiter in weiter Untergruppen aufzuteilen (clustern), wobei deren jeweils konsistente Parameter die Regeln darstellen. Dabei könnte man sowohl agglomerativ (bottom up) als auch divisiv (top down) vorgehen. Ersteres würde bedeuten, dass ausgehend von den jeweiligen Preisänderungen im einzelnen diejenigen mit der höchsten Konsistenz zusammen genommen werden, bis weiteres Zusammenfügen zu viele, zu inkonsistente Parameter in den Gruppen ergeben würde. Die Alternative wäre den Datensatz, wie bei einer Baumstruktur, durch immer feiner werdende Unterscheidungen weiter aufzuspalten, bis vernünftige Regeln entstehen. Die Gruppe mit den Daten, auf die nicht reagiert wurde, wird auch weiter untersucht, weil Muster in ausbleibenden Reaktionen natürlich auch Rückschlüsse über das Konkurrenzverhalten zulassen.

4.5.3 Klassifikation Konkurrenz

4.6 Automatisierungsgrad bestimmen

Wenn man nach dem oben beschriebenen Tool geht, gibt es drei Stufen der Automatisierung. Tankstellen können keines dieser Tools verwenden und auf andere Mittel zurückgreifen. Sie können sich vom Tool Preisänderungen Vorschlägen lassen, diese jedoch manuell nach

eigenem Ermessen durchführen. Oder sie können das Pricing vollautomatisch durch diese Tools betreiben lassen. Die letzten beiden Methoden können natürlich wieder zeitlich variabel abgewechselt werden. Um diese drei Varianten unterscheiden zu können, müssen die unterschiedlichen Reaktionen aufs Pricing Verhalten analysiert werden. Die höchste Stufe der Automatisierung zeichnet sich dadurch aus, dass die in diesem Zeitraum liegenden Regeln keine Ausnahmen haben und die Varianz der Reaktionszeit äußerst gering ist, da hier automatisch und regelbasiert agiert wird. Bei der nächstniedrigeren Stufe, dem Folgen von Vorschlägen, sollten nur geringfügig Abweichungen zum automatischen Betrieb auftauchen, und zwar in der Form, dass Regeln ab und zu verletzt werden und die Varianz der Reaktionszeit merklich ansteigt. Beim arbeiten ohne diese Tools, kann zwar nicht sicher gesagt werden, nach welchen Kriterien hier gearbeitet wird, es sollte jedoch anzunehmen sein, dass das Verhalten weniger regelkonform ist und mit höheren Reaktionszeiten einhergeht.

4.6.1 Formen der Automatisierung

4.6.2 Regelverletzung und Reaktionszeiten

5 Ergebnisse

5.1 Beschreibung

5.2 Diskussion

5.2.1 Statistiken

5.2.2 Problemanalyse

5.3 Zusammenfassung

6 Fazit

6.1 Weitere Möglichkeiten

6.1.1 Gegenrichtung prüfen

6.1.2 Reaktionen an Regeln trimmen

6.1.3 Auslöser untereinander vergleichen

6.1.4 Constraint Satisfaction