

FHNW Transformation Day

Automated Machine Learning und Künstliche Intelligenz mit Datarobot

02.09.2020

Oliver Ruf
Ayse Alemdar
Purushoththaman Shanmuganathan

WER SIND WIR?



Oliver Ruf

☎ +41 61 826 80 80

✉ rvo@informattec.com

MSc in Business Information Systems
Head of Consulting & Projects
Mitglied der Geschäftsleitung | Partner



Ayse Alemdar

☎ +41 61 826 80 80

✉ ala@informattec.com

BSc in Wirtschaftsinformatik
BI Consultant



Purushoththaman Shanmuganathan

☎ +41 61 826 80 80

✉ shp@informattec.com

MSc in Business Information Systems
BI Consultant

1. Übersicht Data Science und Predictive Analytics (20')

2. Gruppenarbeit Data Profiling «Lending Club» (15')

3. Prediction «Lending Club» Daten mit Data Robot und Qlik Sense (20')

4. Fragen (5')

UNSERE KPI's



1

BI-Boutique



4

Basel /
Zürich / Frankfurt / Lissabon



21

Jahre Erfahrung



34

Qualifizierte Mitarbeiter



240+

Zufriedene Kunden



550+

Erfolgreiche Projekte



600+

Training Stunden



10000+

Lizenzierte Anwender

Data
Analytics

Business
Planning

Machine
Learning



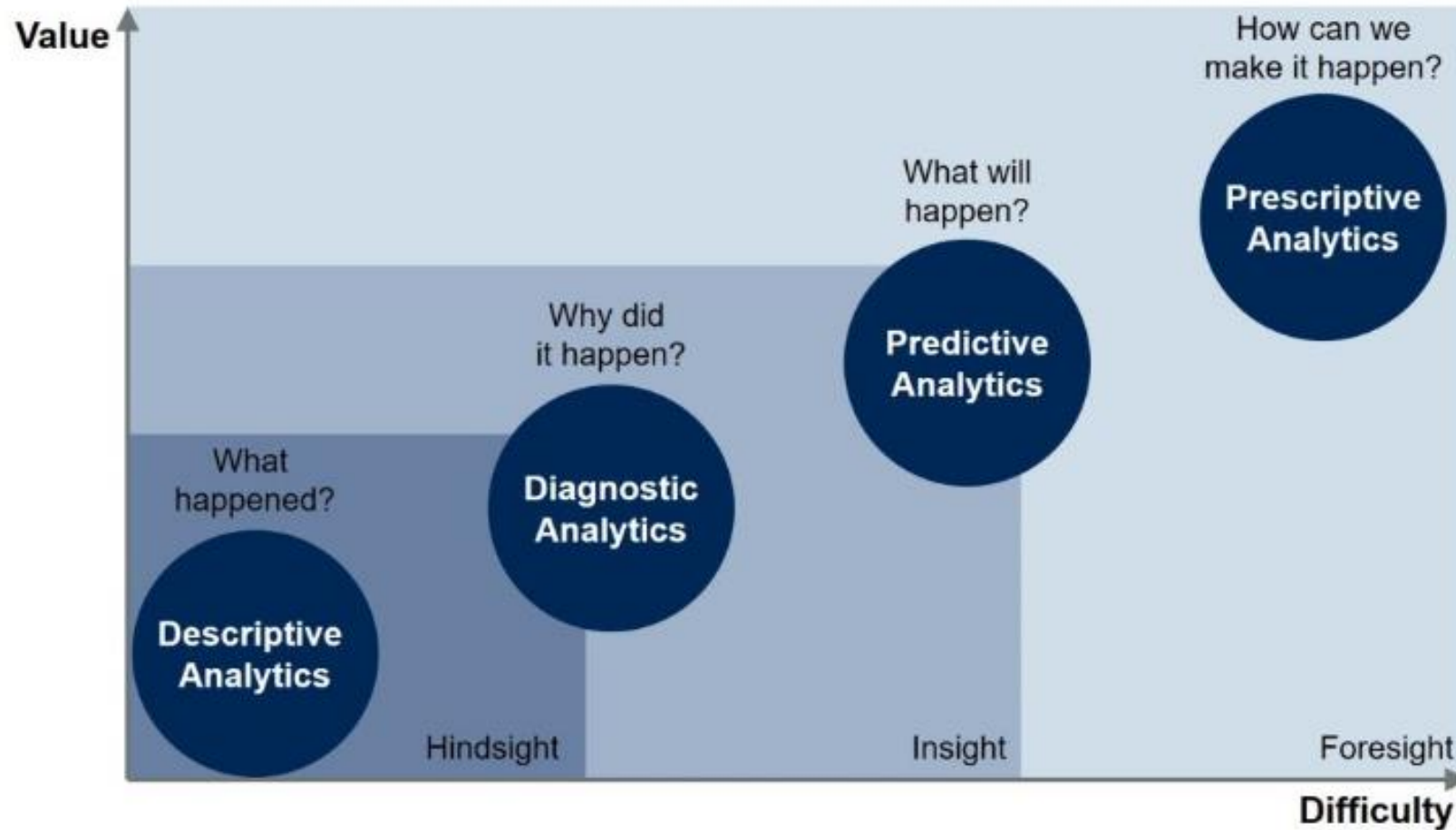
Data Integration



KUNDENAUSZUG



LEVELS OF DATA ANALYTICS



LEVELS OF DATA ANALYTICS (2)



Vorhandene Daten aufbereiten und auswerten

Bsp.: «Wir hatten im letzten Quartal einen Kundenzuwachs von 4% gegenüber dem Vorjahr.»

Lernen aus den vergangenen Daten oder anderen Quellen

Bsp.: «Wir werden einen Kundenzuwachs von 2% im nächsten Quartal haben»

Macht Handlungsvorschläge auf Basis unserer Vorhersage

Bsp.: «Wir werden einen Kundenzuwachs von 2% haben. Deswegen sollten wir für diese Filiale eine MitarbeiterIn zusätzlich einstellen.»

DATA MINING

Das Erkennen von Mustern und Strukturen in Daten.

«What knowledge can I gain from this data?»



Target wusste vor dem Vater, dass Tochter schwanger ist

MACHINE LEARNING

Machine Learning ist ein Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung. Dazu bauen Algorithmen ein statistisches Modell auf, das auf strukturierten Trainingsdaten beruht.

DEEP LEARNING

Deep Learning ist ein Teilbereich von Machine Learning, welcher durch neuronale Netze für das Lernen aus unstrukturierten Daten geeignet ist. Deep Learning kommt im Rahmen künstlicher Intelligenz häufig für die Gesichts-, Objekt- oder Spracherkennung zum Einsatz.



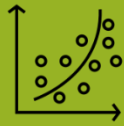


Classification

Classification - Zuordnung von Elementen zu Gruppen:

Wird Kunde A sein Abo in den nächsten 6 Monaten kündigen?

→ Gruppen: Ja / Nein



Regression

Regression - Abschätzung eines Werts:

Wieviele GB an mobilen Daten wird Kunde A in den nächsten 6 Monaten nutzen?

→ Zahl N



Clustering

Clustering - Gruppierung nach gleichzeitigem Auftreten:

Welche Produkte werden häufig zusammen gekauft?

→ Gruppen: A, B, C etc.

VORGEHEN IM DATA SCIENCE PROJEKT

Aufgabenverständnis

- Definition Ziel / Business Nutzen
- Formulierung Zielvariable

Datenvorbereitung & -aufbereitung

- Beschaffung Trainingsdaten
- Datenharmonisierung / Transformation

Modellbildung & -anwendung

- Evaluierung Data Mining Verfahren
- Auswahl oder Entwicklung Modelle
- Training

Validierung, Interpretation & Integration

- Bestätigung des Modells
- Deployment in BI oder operative Systeme

Monitoring / Lifecycle





Healthcare

- *Planung Bettenauslastung*
- *Vorhersage ambulanter Fälle*
- *Reduzieren von Wiederkehrer*



Banking / Finance

- *Risikobeurteilung*
- *Kreditwürdigkeit*
- *Einblicke in Börsenhandel*

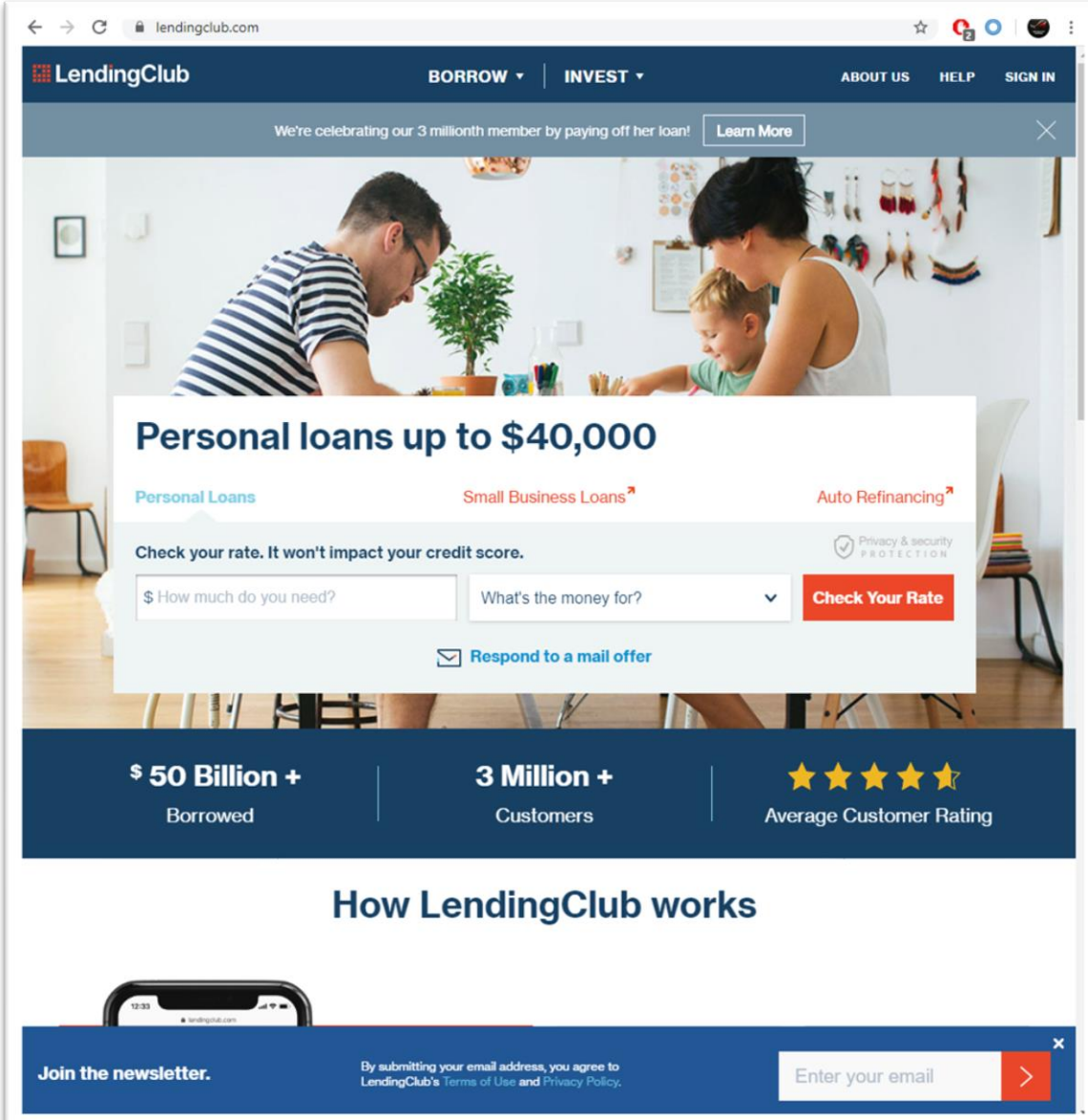


Retail / Sales

- *Risiko erkennen*
- *Kundenklassifizierung*
- *Umsatz vorhersehen*

2. GRUPPENARBEIT DATA PROFILING «LENDING CLUB»

- Datenbeispiel Lending Club
 - Plattform zur Vermittlung von Peer-to-Peer-Krediten
 - Kreditnehmer:
 - Höhere Chance auf Finanzierung
 - Überdurchschnittlich hohe Zinsen
 - Investoren:
 - Höhere Rendite
 - Überdurchschnittlich hohes Risiko

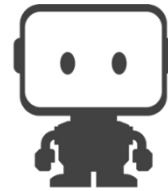


The screenshot shows the LendingClub website interface. At the top, there is a navigation bar with 'BORROW' and 'INVEST' options, along with 'ABOUT US', 'HELP', and 'SIGN IN'. A banner at the top right celebrates a 3 millionth member. The main content area features a large image of a family and a prominent offer for 'Personal loans up to \$40,000'. Below this, there are three tabs: 'Personal Loans', 'Small Business Loans', and 'Auto Refinancing'. A form prompts users to 'Check your rate. It won't impact your credit score.' with input fields for '\$ How much do you need?' and 'What's the money for?', and a 'Check Your Rate' button. A 'Respond to a mail offer' link is also present. At the bottom, three statistics are displayed: '\$ 50 Billion + Borrowed', '3 Million + Customers', and 'Average Customer Rating' (5 stars). A footer section titled 'How LendingClub works' includes a newsletter sign-up form with the text 'Join the newsletter.' and 'By submitting your email address, you agree to LendingClub's Terms of Use and Privacy Policy.'

2. GRUPPENARBEIT DATA PROFILING «LENDING CLUB»

- Anonymisierte Echtdaten
 - 30 Records aus einem Dataset von 10'000 Einträgen
 - 30 Features (Felder)
- Zielvariable für ein Data Science Projekt:
 - «is_bad» [0|1]
 - Wie hoch ist die Wahrscheinlichkeit, dass «is_bad» 1 ist?
 - Sprich, **Ausfallwahrscheinlichkeit des Kreditnehmers**
- Aufgabe:
 - Welche Einflussfaktoren auf unsere Zielvariable sehen Sie? Ordnen Sie die 5 wichtigsten Features nach Relevanz (1 = Wichtigstes).
 - Was sind die Challenges in den Daten?

3. PREDICTION «LENDING CLUB» DATEN MIT DATA ROBOT



DataRobot



ERLEBEN SIE DATAROBOT LIVE...

QLIK DISCOVERY & BOOSTER DAY 2020



24.09.2020

Ambassador House Zürich

HIGHLIGHTS:

- Kundenvorträge
Migros Bank, Stäubli & Ronal Group
- Qlik Hands-on Workshops
- Breakout Sessions
(Forecasting / Augmented Insights)
- Jedox Live Demo
- **AI & AutoML mit DataRobot**
- Partner Expo
- ...und vieles mehr

#iQDBD2020

kostenlose Anmeldung: www.informattec.com/iQDBD2020