

RECEIPT DATA EXTRACTION

BY TAMINO, DIEGO AND JAMIAN

FINAL PRESENTATION

DSPRO2

GOAL



```
{
  "Shopname": "Migros",
  "Date and time": "15.03.2025 19:47",
  "Products": [
    {
      "name": "Hummus Piccante 200g",
      "amount": 1,
      "price": "3.80",
      "discount": "0.00"
    },
    {
      "name": "MBud Eisbergsalat",
      "amount": 1,
      "price": "2.00",
      "discount": "0.00"
    },
    {
      "name": "Ferrero Rocher",
      "amount": 1,
      "price": "17.85",
      "discount": null
    },
    {
      "name": "Plastiksack",
      "amount": 2,
      "price": "0.05",
      "discount": null
    }
  ],
  "Tax": {
    "2.6%": "0.60",
    "8.1%": "0.00"
  },
  "Total price": "23.70"
}
```

GETTING DATA



time consuming



expensive



hard

OUR SOLUTION

SYNTHETIC DATA

Shop name: Migros

Date: 05.09.2017

Time: 15:42

Products:

Lager Bier: 2x, 9.45

Avocado: 6x, 3.95

...

Migros

Luzern Löwencenter
Tel. +41 15 677 29 98
Migros
Mattweg 41, 6925 Luzern

Artikel	Menge	Preis	Aktion	Total
Wachteleier	5	0.75		3.75 0
Äpfel Jazz	1	19.80		19.80 0
Rindsschnitzel	1	9.35		9.35 0
Premium Beerenmischung	1	1.60		1.60 0

Total CHF 34.50

Debit Mastercard contactless
#XXXXXXXXXXXX0860/23.10.2021/04:00:26/56882282/*36
5830*32635989692/1A2B3C/32635989692#
Total-EFT CHF: 34.50

MIGROS, CHE-750.954.473 MWST			
GR	MWST%	TOTAL	MWST
0	8.0	34.50	2.76

23.10.21 04:00 12345 12345678 01 1234567





Migros

Luzern Löwencenter
Tel. +41 15 677 29 98
Migros
Mattweg 41, 6925 Luzern

Artikel	Menge	Preis	Aktion	Total
Wachteleier	5	0.75		3.75 0
Äpfel Jazz	1	19.80		19.80 0
Rindsschnitzel	1	9.35		9.35 0
Premium Beerenmischung	1	1.60		1.60 0

Total CHF 34.50

Mastercard 34.50
Total in EUR 32.34

Debit Mastercard contactless
#XXXXXXXXXXXX0860/23.10.2021/04:00:26/56882282/*36
5838*32635989692/1A2B3C/32635989692#
Total-EFT CHF: 34.50

MIGROS, CHE-750.954.473 MWST
GR MWST% TOTAL MWST
0 8.0 34.50 2.76

23.10.21 04:00 12345 12345678 01 1234567

Migros

Luzern Löwencenter
Tel. +41 15 677 29 98
Migros
Mattweg 41, 6925 Luzern

Artikel	Menge	Preis	Aktion	Total
Wachteleier	5	0.75		3.75 0
Äpfel Jazz	1	19.80		19.80 0
Rindsschnitzel	1	9.35		9.35 0
Premium Beerenmischung	1	1.60		1.60 0

Total CHF 34.50

Mastercard 34.50
Total in EUR 32.34

Debit Mastercard contactless
#XXXXXXXXXXXX0860/23.10.2021/04:00:26/56882282/*36
5838*32635989692/1A2B3C/32635989692#
Total-EFT CHF: 34.50

MIGROS, CHE-750.954.473 MWST
GR MWST% TOTAL MWST
0 8.0 34.50 2.76

23.10.21 04:00 12345 12345678 01 1234567



SPAR				
Luzern Lebensmittel Tel. 041 85 942 42 76 Biere Bahnhofstrasse 1, 6007 Bern				
Artikel	Menge	Preis	Aktion	Total
Sonnenblumenöl	2,707	56.90		154.88 0
Beislakun	1	5.30		18.68 0
Brustfleisch	1	4.85		4.85 0
Pommes	1	8.80		8.80 0
Butterbrotling	1	26.55		26.55 0
Kuchentopf	1	25.80		25.80 0
Eisbergsalat	1	5.90		13.88 0
Meisen	1	6.15		6.15 0
Cashew	2	4.85		9.70 0
Tafelberg				
Total CHF				199.55
Netto				199.55
Total in EUR				181.42
Debit Mastercard contactless XXXXXXXXXXXX 0000 0000 0000 0000 6011 0000 0000 0000 0000 0000 0000 0000 Total-EFT CHF: 199.55				
SPAR, CHE-519 454.312 MIST GR MIST TOTAL 0 4.3 199.55 6.50				
89-20-22 98-12 12345 12345678 91 1234567				

Netto				
Luzern Lebensmittel Tel. 041 85 942 42 76 Netto Bahnhofstrasse 1, 6007 Bern				
Artikel	Menge	Preis	Aktion	Total
Zahn	2	8.30		16.60 0
Beislakun	1	5.30		18.68 0
Brustfleisch	1	4.85		4.85 0
Pommes	1	8.80		8.80 0
Butterbrotling	1	26.55		26.55 0
Kuchentopf	1	25.80		25.80 0
Eisbergsalat	1	5.90		13.88 0
Meisen	1	6.15		6.15 0
Cashew	2	4.85		9.70 0
Tafelberg				
Total CHF				123.75
Netto				123.75
Total in EUR				118.45
Debit Mastercard contactless XXXXXXXXXXXX 0000 0000 0000 0000 6011 0000 0000 0000 0000 0000 0000 0000 Total-EFT CHF: 123.75				
REWE, CHE-519 454.312 MIST GR MIST TOTAL 0 4.3 123.75 6.50				
89-20-22 98-12 12345 12345678 91 1234567				

ALDI				
Luzern Lebensmittel Tel. 041 85 942 42 76 ALDI Bahnhofstrasse 1, 6007 Bern				
Artikel	Menge	Preis	Aktion	Total
Curry schief	1	18.30		18.30 0
Curry	2	5.45		10.90 0
Sofra	2	5.40		10.80 0
Sofra Williams 8.5 g	1	8.80		8.80 0
Wurst	1	5.30		18.68 0
Brustfleisch	1	4.85		4.85 0
Budget Yogurt	1	18.68		18.68 0
Croissant	1	1.30		1.30 0
Langkornreis	1	1.30		1.30 0
Zitrus	1	1.30		1.30 0
Cherrytomaten	1	1.30		1.30 0
Schokoladefleisch	1	1.30		1.30 0
Fasili 500 l	1	1.30		1.30 0
Total CHF				495.10
Netto				495.10
Total in EUR				451.85
Debit Mastercard contactless XXXXXXXXXXXX 0000 0000 0000 0000 6011 0000 0000 0000 0000 0000 0000 0000 Total-EFT CHF: 495.10				
ALDI, CHE-519 454.312 MIST GR MIST TOTAL 0 4.3 495.10 6.50				
89-20-22 98-12 12345 12345678 91 1234567				

REWE				
Luzern Lebensmittel Tel. 041 85 94 26 96 REWE Bahnhofstrasse 1, 6007 Luzern				
Artikel	Menge	Preis	Aktion	Total
Wasserschwein	1	17.50		17.50 0
Fasili	1	1.50		1.50 0
Küchenpfanne	1	3.15		3.15 0
Wasser Pfeffer	1	8.85		8.85 0
Küchenrolle 600 g	1	2.20	2.85	28.25 0
Brot 100g				8.15-
Schinken	1	21.35		21.35 0
Budget Braten	3	31.50		46.50 0
Wurst-Linsen	1	38.30		38.30 0
Curry	2	8.30		33.20 0
Chokolade	9	38.25		32.25 0
Meisen	2	5.80		33.80 0
Wurst-Linsen	1	38.30		38.30 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst	1	3.20		3.20 0
Wurst				

PREPROCESSING



Netto

Marken-Discount

Heimberg Shopping

Artikel	Menge	Preis	Aktion	Total
Honigmelone	1	13.00		13.00 0
Safran	1	17.90		17.90 0
Budget Brezel gross	3	15.50		46.50 0
Teller-Linsen	3	10.20		30.60 0
Sonnenblumenöl	4	8.80		35.20 0
Gurken	2	8.20		16.40 0

Total CHF

159.60

Mastercard

159.60

Total in EUR

165.05

Debit

Mastercard contactless

#XXXXXXXXXXXX9105/23.03.1996/04:24/10209757/*65211

3*71815614175/1A2B3C/71815614175#

Total-EFT CHF:

159.60

NETTO, CHE-626.343.385 MWST

GR MWST% TOTAL MWST

0 3.6 159.60 5.75

Danke vell mol för Ere Ichauf

5386237168334581402154436842347119

Filiale Bedien. Nbr Bcn Datum Zeit

5104937 5104937 556 0181 23.03.96 04:24:18 C

Netto

Marken-Discount

Heimberg Shopping

Artikel	Menge	Preis	Aktion	Total
Honigmelone	1	13.00		13.00 0
Safran	1	17.90		17.90 0
Budget Brezel gross	3	15.50		46.50 0
Teller-Linsen	3	10.20		30.60 0
Sonnenblumenöl	4	8.80		35.20 0
Gurken	2	8.20		16.40 0

Total CHF

159.60

Mastercard

159.60

Total in EUR

165.05

Debit

Mastercard contactless

#XXXXXXXXXXXX9105/23.03.1996/04:24/10209757/*65211

3*71815614175/1A2B3C/71815614175#

Total-EFT CHF:

159.60

NETTO, CHE-626.343.385 MWST

GR MWST% TOTAL MWST

0 3.6 159.60 5.75

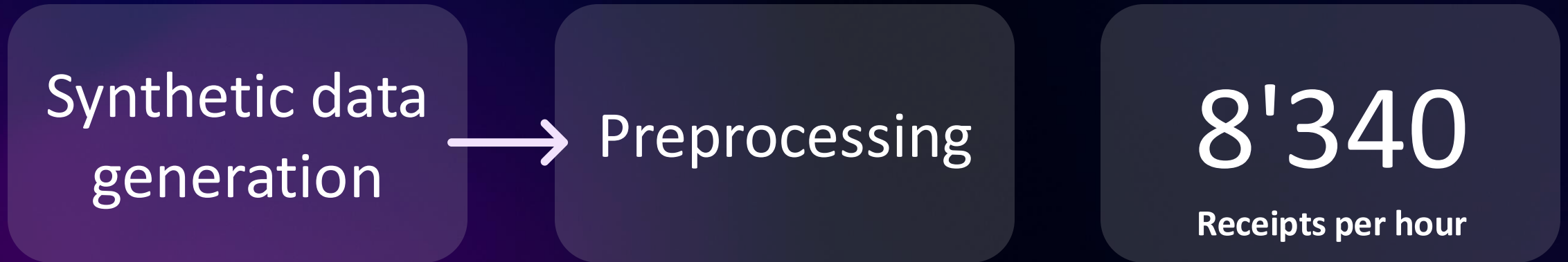
Danke vell mol för Ere Ichauf

5386237168334581402154436842347119

Filiale Bedien. Nbr Bcn Datum Zeit

5104937 5104937 556 0181 23.03.96 04:24:18 C

DATA GENERATION



DATASET COMPOSITION

Train & validation
Synthetic data

Test #1
Synthetic data

Test #2
Coop & Migros

Test #3
Other Grocery
Stores

Test #4
Different
domains

MODEL



DONUT

OCR FREE

201 M

document understanding transformer

VISION ENCODER

swin



LANGUAGE DECODER

bart



question answering



classification



extraction

TASK TYPE

DEMO

evaluation metric

LEVINSTEIN

- **delete** a character
- **update** a character
- **create** a character

operations

cat -> bat
replace 'c' with 'b'
= levinstein distance 1

example

test metric

TREE EDIT DISTANCE

- **delete** a node
- **update** a node
- **create** a node

operations



= 3

example

RESULTS

99 %

test #1

80 %

test #2

31 %

test #3

8 %

test #4

test #1 - synthetic data

test #2 - real data: migros & coop

test #3 - real data: different grocery stores (edeka & denner)

test #4 - real data: different domains (book store, gas station, ...)

RTX 3090

NVIDIA GPU

24

vRAM

29

RAM

29

GB storage

9

vCPU

runpod

serverless cloud provder

CLOUD HARDWARE

23 min

duration

3

batch-size

3

epochs

3e-5

learning rate

1290x960

input dimension

1'000

train size

200

valid size

TRAINING FACTS

0 %

donut base

28 %

smoIVLM

80 %

our solution

test set #2

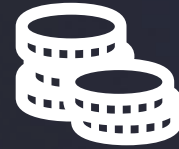
COMPARISON

USECASE

API that can be integrated into various software products in which people work with receipts



accounting



Expense management

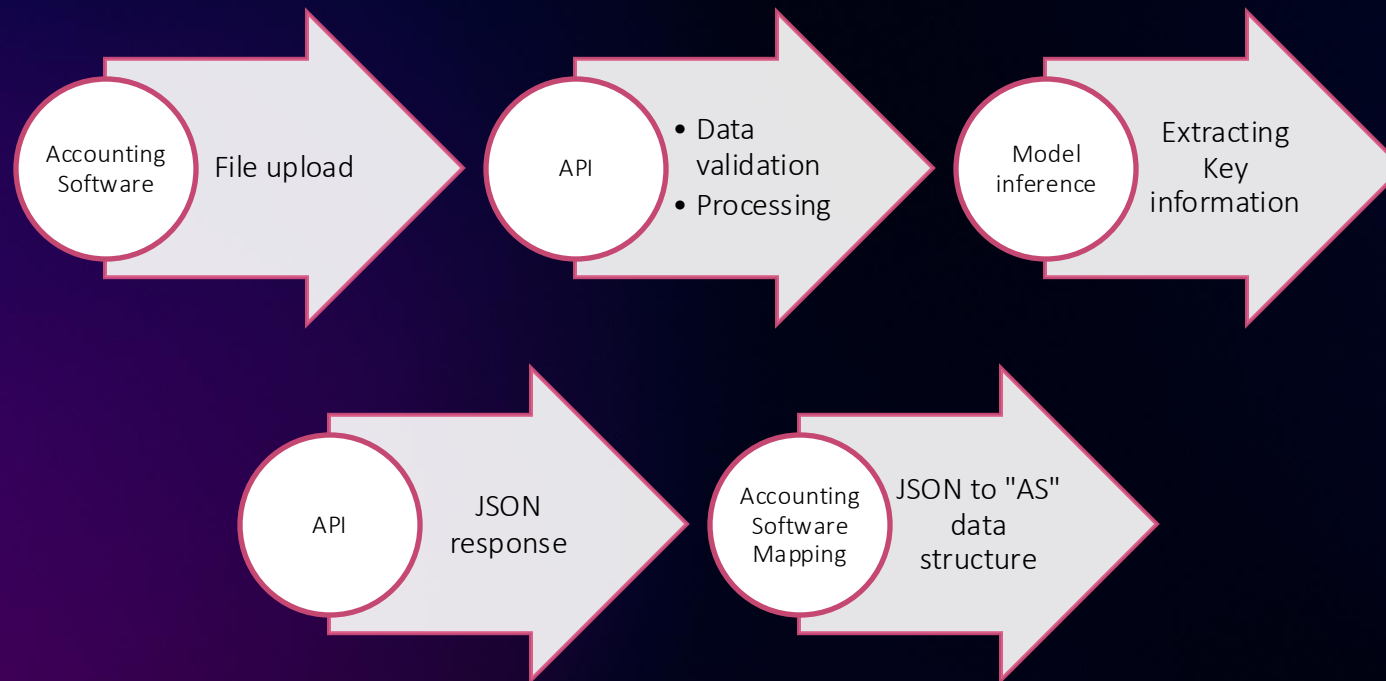


tax



personal finance

USECASE: ACCOUNTING





fast



privacy



less CO₂



saving



accurate

CHOOSE US

POSSIBLE IMPROVEMENTS

- Improve receipt generation
 - More diverse templates
 - More possible products & placeholder values
- Develop complete custom model for this task

THANK YOU

Q&A SESSION