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
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
#XXXXXXXXXXX8860/23.10.2021/04:00:26/56882282/*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
#XXXXXXXXXXX0860/23.10.2021/04:00:25/56882282/*36
5838*32635939692/1A2B3C/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
#XXXXXXXXXXX0860/23.10.2021/04:02:55/56882282/*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 04:02 12345 12345678 01 1234567

SPAR
Total CHF
199.55

Netto
Marken-Discount
Total CHF
123.75

ALDI
Total CHF
495.10

REWE
Total CHF
476.70

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.	KNr	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.	KNr	Bcn	Datum	Zeit
5104937	5104937	556	0181	23.03.96	04:24:18 C

DATA GENERATION

Synthetic data generation → Preprocessing

8'340

Receipts per hour

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

Tree A:

Tree B:

= 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, ...)

CLOUD HARDWARE

RTX 3090
NVIDIA GPU

24
vRAM

29
RAM

29
GB storage

9
vCPU

runpod
serverless cloud provder

TRAINING FACTS

23 min duration	**3** batch-size	**3** epochs	**3e-5** learning rate
1290x960 input dimension	**1’000** train size	**200** valid size	

COMPARISON

0 %

donut base

28 %

smolVLM

80 %

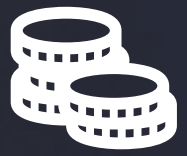
our solution

test set #2

USECASE

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

accounting

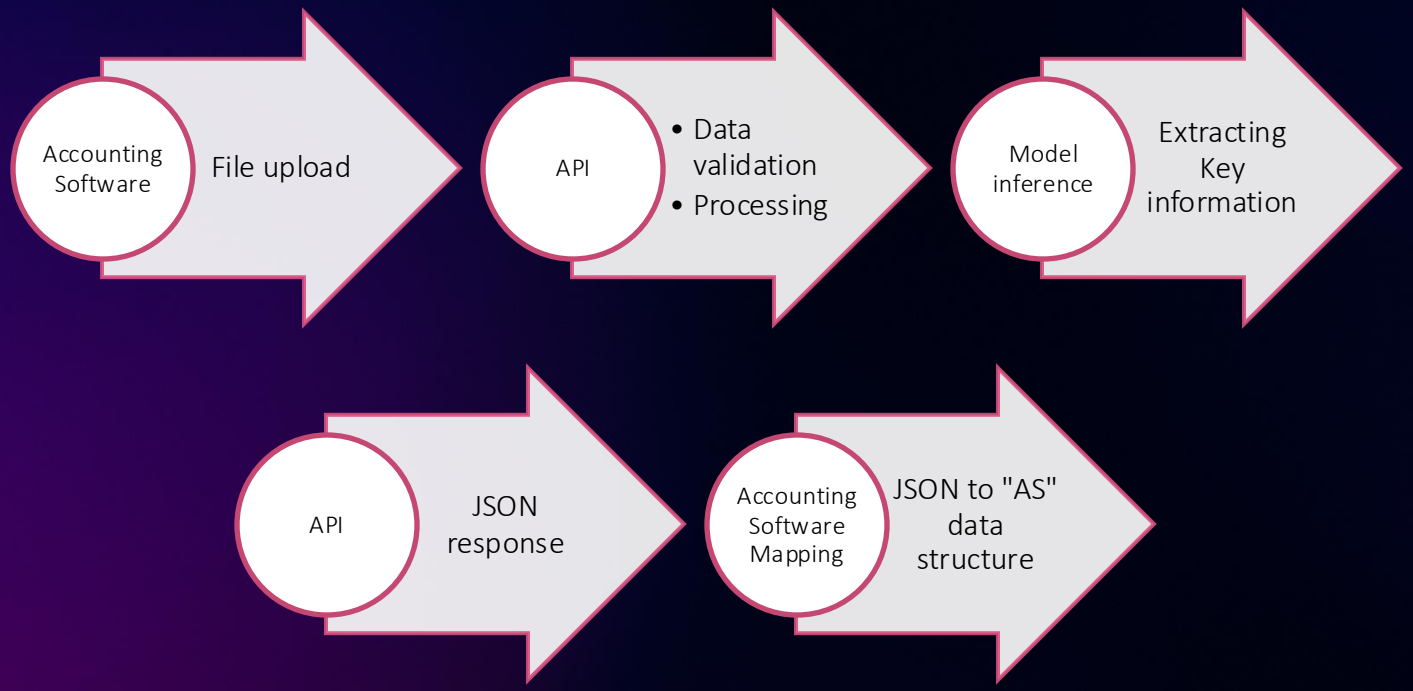


Expense management

tax

personal finance

USECASE: ACCOUNTING



fast

privacy

less co_2

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