

"Initiative Latvia" SIA
Mākslīgā intelekta plānošanas
sistēmas
Tehniskā specifikācija

Saturs

Saturs	2
1. Kopsavilkums	5
1.1. Galvenās iespējas	5
2. Sistēmas arhitektūra	6
2.1. Tehnoloģiskais stacks	6
2.2. Arhitektūras diagramma	7
3. Datu Integrācija	8
3.1. Kantar AdEx Integrācija	8
ETL process	8
Galvenie lauki	8
3.2. Kantar Atlas integrācija	9
Datu struktūra	9
Ielādes process	9
4. Datu noliktavas shēma	9
4.1. Faktu (<i>Fact</i>) tabulas	9
4.2. Dimensiju (<i>Dim</i>) tabulas	10
5. NLP auditoriju kartēšana	12
5.1. Četru soļu <i>Pipeline</i>	12
1. solis: noteikumu balstīts parsētājs (<i>Rule-Based Parser</i> , aptv. 70 % pārklājumu)	12
2. solis: nosaukto entītiju atpazīšana (<i>Named Entity Recognition</i>)	12
3. solis: Semantiskā līdzība (<i>Embeddings + Fuzzy Matching</i>)	12
4. solis: LLM rezerves mehānisms (<i>LLM Fallback</i>)	13
5.2. Cilvēka līdzdalības saskarne (<i>Human-in-the-Loop UI</i>)	13
5.3. Veiktspējas rādītāji un precizitāte	13
6. Optimizācijas dzinējs	15
6.1. Matemātiskais formulējums	15
6.2. Reakcijas līkņu (<i>Response Curve</i>) modelis	15
Treniņa process	16
6.3. Solvera (<i>cvxpy</i>) ieviešana	16
6.4. Atkāpšanās (<i>fallback</i>) mehānisms	17
7. AI plānošana viena medija letvaros	17
7.1. Procesa soļi	17
7.2. Izvada piemērs	18
8. Izskaidrojamība (<i>explainability</i>)	20
8.1. Izskaidrojuma komponentes	20

8.2. LLM stāstījuma <i>pipeline</i>	20
8.3. LLM stāstījuma izvada piemērs	21
9. API Specifikācija	22
9.1. Pamatendpunkti (<i>Core Endpoints</i>)	22
POST /api/v1/campaigns/optimize.....	22
POST /api/v1/audiences/map	23
GET /api/v1/campaigns/{id}/performance	23
9.2. Autentifikācija.....	24
Lietotāju piekļuves kontrole.....	24
Drošības ierobežojumi	24
9.3. Kļūdu atbildes (<i>Error Responses</i>)	24
Formāts	24
Piemēri	24
10. Monitorings un veikspēja	26
10.1. Galvenie rādītāji	26
10.2. Informācijas paneļa komponentes	26
10.3. Brīdinājumu stratēģija	26
Kritiski (SMS + Tālrunis)	26
Augsts (SMS)	26
Vidēja prioritāte (e-pasts)	26
11. Testēšana un validācija	27
11.1. Vienību testi (<i>Unit tests</i>).....	27
11.2. Integrācijas testi (<i>Integration tests</i>).....	27
11.3. Atpakaļtestēšana (<i>Backtesting</i>).....	27
11.4. Zelta datu kopas (<i>Golden Datasets</i>).....	27
12. Ieviešana, tai skaitā pakāpeniska ieviešana	29
12.1. Pakāpeniskas ieviešanas plāns	29
1. fāze – <i>Shadow mode</i> (3 nedēļas)	29
2. fāze – Pilotprojekts (2 mēneši)	29
3. fāze – Paplašināšana (2 mēneši).....	29
4. fāze – Pilnīga automatizācija (3+ mēneši)	29
12.2. Atsaukšanas (rollback) nosacījumi	29
13. Riska faktori un mazināšanas pasākumi	30
14. Panākumu kritēriji	31
14.1. Tehniskie kritēriji (MVP).....	31
14.2. Biznesa kritēriji	31
15. Nākotnes uzlabojumi	32
15.1. Versija v1.1 (plānots 2026. gada 3. ceturksnis)	32
15.2. Versija v1.2 (plānots 2026. gada 4. ceturksnis)	32
15.3. Versija v2.0 (plānots 2027. gada 1. ceturksnis)	32

16. Pielikumi	33
16.1. String → Kantar Atlas kartēšanas piemēri	33
16.2. Atlaižu liknes piemērs	33
TV atlaižu struktūra (SCD2)	33
17. Terminu vārdnīca	34

1. Kopsavilkums

Šī tehniskā specifikācija definē pilnīgu AI balstītu mediju plānošanas sistēmu, kas automātiski optimizē mediju sadalījumu, balstoties uz klienta KPI mērķiem un biznesa ierobežojumiem.

1.1. Galvenās iespējas

- ETL konveijers Kantar AdEx (MS SQL) un Kantar Atlas (CSV) datu integrācijai
- NLP vadīta auditoriju kartēšana ar cilvēka validācijas posmu
- Optimizācijas dzinējs ar komerciālo nosacījumu atbalstu
- Reakcijas likņu modelis (*response curve*) KPI prognozēšanai
- AI vadīta plānošana viena medija ietvaros (kanāli, programmas, laika joslas)
- LLM vadīta izskaidrojamība un audita žurnāls

Sistēma nodrošina vismaz 10% augstāku efektivitāti salīdzinājumā ar manuālo plānošanu, saglabājot pilnīgu pārredzamību un iespēju reproducēt rezultātu.

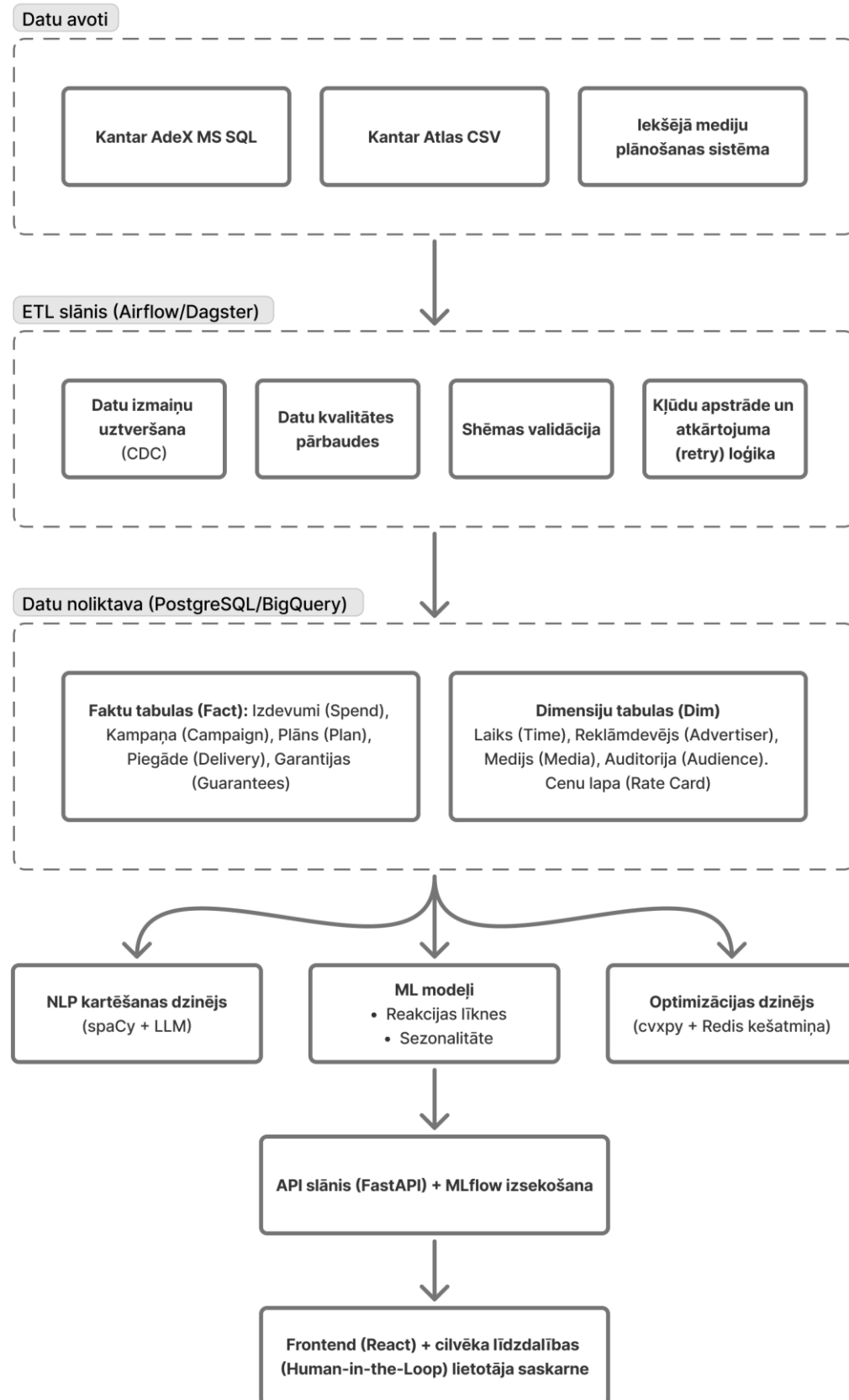
2. Sistēmas arhitektūra

2.1. Tehnoloģiskais stacks

Slānis	Tehnoloģija	Pamatojums
ETL / orķestrācija	Nav noteikta	Uzticams darba plūsmu (<i>workflow</i>) vadības risinājums ar atkārtojuma (<i>retry</i>) loģiku un brīdinājumu (<i>alerting</i>) mehānismu
Transformācijas	Nav noteikta	SQL transformācijas ar versiju kontroli un testēšanu
Datu noliktava (Data Warehouse)	PostgreSQL 15+ vai Google BigQuery	Mērogojams OLAP risinājums ar partīciju atbalstu
Optimizācija	Python 3.11+ (cvxpy 1.4+, pulp 2.7+)	Konveksās optimizācijas bibliotēkas ar komerciālu atbalstu
ML Pipeline	scikit-learn 1.4+, statsmodels 0.14+	Reakcijas līkņu (<i>response curve</i>) pielāgošana un sezonālītātes modelēšana
ML Ops	Nav noteikta	Modeļu versiju kontrole un reproducējamība
NLP/LLM	sentence-transformers, spaCy, LLM	Auditoriju kartēšana un izskaidrojamības ģenerēšana
API	FastAPI 0.110+	Asinhrons API ar automātisku dokumentāciju (OpenAPI)
Caching	Redis 7.2+	Optimizācijas rezultātu un auditoriju kartējumu kešošana
Frontend	React 18+	Moderna vienlapas lietotne (SPA) ar komponentu atkārtotu izmantošanu
Uzraudzība	Prometheus + Grafana	Metriku vākšana, uzraudzība un vizualizācija

2.2. Arhitektūras diagramma

Sistēma sastāv no šādām galvenajām komponentēm:



3. Datu Integrācija

3.1. Kantar AdEx Integrācija

Datu avots: Microsoft SQL Server datubāze ar dienas/nedēļas atjauninājumiem

ETL process

- Savienojums un datu ieguve (*Connection & Extraction*)
 - Izmantot SQLAlchemy ar pyodbc draiveri
 - Inkrementālā ielāde: `WHERE modified_date > last_load_timestamp`
 - Savienojumu baseins (*Database Connection Pool*): `min = 5, max = 20, timeout = 30 s`
- Datu kvalitātes pārbaudes (*Data Quality Checks*)
 - `NULL` pārbaude kritiskajos laukos (pieļaujamā novirze < 1 %)
 - Datumu diapazona validācija: ± 7 dienas no aktuālā datuma
 - Valūtas konsekvence: visām vērtībām jābūt EUR
 - Dublikātu noteikšana: `UPSERT` ar konfliktu atrisināšanu
 - Izmaksu validācija: `gross  $\geq$  net, discount  $\in$  [0, 100]`
- Transformācijas (*Transformations*)
 - Valūtas konvertācija uz EUR (ja nepieciešams)
 - PVN korekcijas atkarībā no klienta tipa
 - GRP normalizācija (15 minūšu konsolidācija)
 - Datuma/laika standartizācija (UTC)
- Kļūdu apstrāde (*Error Handling*)
 - Atkārtojuma loģika: 3 mēģinājumi ar eksponenciālu aizturi (2 s, 4 s, 8 s)
 - Drošības slēdzis (*Circuit Breaker*): apturēt pēc 5 secīgiem kļūmīgiem mēģinājumiem
 - Bīdinājumi: e-pasts pie kļūdām
 - Nederīgo ierakstu karantīna pārskatīšanai (*Dead Letter Queue*)

Galvenie lauki

Lauks	Tips	Apraksts
Medium	VARCHAR(50)	TV, radio, internets (web), prese vai vides reklāma (OOH)
Channel	VARCHAR(100)	Piemēram, TV3, LTV1, Delfi u.c.
Cost_EURO	DECIMAL(12,2)	Bruto vai neto vērtība, atkarībā no aģentūras uzskaites principa

Lauks	Tips	Apraksts
GRP	DECIMAL(8,2)	Auditorijas reitings vai kopējais sasniedzamības rādītājs
Discount_INT	DECIMAL(5,2)	Piemērotais atlaides apjoms procentos
Date	DATE	Reklāmas rādīšanas vai kampaņas aktivitātes datums

3.2. Kantar Atlas integrācija

Datu avots: CSV fails (1× gadā, ~Q1)

Datu struktūra

- **Auditoriju segmenti:** dzimums, vecums, reģions, ienākumu līmenis, izglītība, intereses
- **Metrikas:** Reach, Frequency, Affinity Index, Size (tūkstoši)
- **Ģeogrāfija:** valsts līmenis (LV, EE, LT) un reģioni (Rīga, Pierīga, Kurzeme u.c.)

Ielādes process

1. CSV validācija

- **Shēmas pārbaude:** paredzētās kolonnas un datu tipi
- **Kodējums:** UTF-8 ar BOM
- **Atdalītāja noteikšana:** komats / semikols / tabulators

2. Starpposma ielāde (Staging)

- Ielādēt neapstrādāto CSV → [Staging.AtlasRaw](#)
- Pievienot metadatus: [file_name](#), [load_timestamp](#), [row_hash](#)

3. Normalizācija

- Parsēt auditoriju atribūtus → [DimAudience](#) (SCD Type 2)
- Izveidot hierarhisku taksonomiju: **Dzimums** → **Vecums** → **Reģions** → **Interese**
- Ielādēt metriku → FactAudienceMetrics ar ārējo atslēgu uz [DimAudience](#)

4. Versiju kontrole

- **SCD Type 2 lauki:** [valid_from](#), [valid_to](#), [is_current](#)
- Iepriekšējās versijas deaktivēšana ([is_current](#) = [false](#)) jaunās ielādes brīdī
- Lietotāju paziņojums: “*Jaunie Atlas dati (2026) ir pieejami*”

4. Datu noliktavas shēma

4.1. Faktu (*Fact*) tabulas

Tabula	Mērķis un galvenie lauki
FactSpend	Vēsturiskie izdevumi no AdEx date_key, advertiser_key, medium_key, channel, cost_eur, grp, discount_pct
FactCampaign	Kampaņu metadati campaign_id, advertiser_key, start_date, end_date, budget_total, kpi_target
FactPlan	AI ģenerētie plāni plan_id, campaign_id, medium_key, channel, spend_planned, grp_expected, version
FactDelivery	Faktiskie rezultāti plan_id, medium_key, channel, spend_actual, grp_actual, variance_pct
FactGuarantees	Garantētie apjomi vendor, medium, min_spend_eur, max_discount_pct, valid_from, valid_to
FactFees	Aģentūras atlīdzības modelis advertiser_key, fee_type (pct/fixed/hybrid), fee_value, rebate_pct
FactPlan_MediaLevel	Viena medija plāni plan_id, medium, channel, program, time_band, spend, grp_expected, prime_ratio

4.2. Dimensiju (Dim) tabulas

Dimensija	Atribūti un SCD tips
DimTime	date, year, quarter, month, week, weekday, is_holiday SCD Type 1
DimAdvertiser	advertiser_id, name, industry, country SCD Type 1
DimMedia	medium, channel, owner, rate_card_link, valid_from, valid_to SCD Type 2
DimAudience	audience_id, gender, age_min, age_max, region, interests, size_k, valid_from, valid_to SCD Type 2

Dimensija	Atribūti un SCD tips
DimMediaSlot	channel, time_band, prime_flag, avg_grp, avg_cost, audience_affinity SCD Type 1

5. NLP auditoriju kartēšana

Mērķis: automātiski pārveidot brīvā teksta formā uzrakstītas auditorijas specifikācijas strukturētā `DimAudience` ierakstā, nodrošinot cilvēka validācijas posmu. Šis posms ir rekomendēts, bet ne obligāts.

5.1. Četru soļu Pipeline

1. solis: noteikumu balstīts parsētājs (*Rule-Based Parser*, aptv. 70 % pārklājumu)

Izmantot **regex** veidnes vienkāršiem rakstu (*pattern*) atpazīšanas gadījumiem:

```
GENDER_PATTERN = r'(vīrieši|sievietes|men|women|males|females)'
AGE_PATTERN = r'(\d{2})[--](\d{2})' # Match 18-45
REGION_PATTERN = r'(Rīga|Pierīga|Kurzeme|Vidzeme|Zemgale|Latgale)'
```

Piemērs:

```
Input: 'Sievietes 20-50, Rīga un Pierīga'
Output: {gender: 'F', age: [20, 50], regions: ['Rīga', 'Pierīga']}
Confidence: 0.95
```

2. solis: nosaukto entītiju atpazīšana (*Named Entity Recognition*)

Izmantot **spaCy NER** modeli, apmācītu uz pielāgotiem datiem:

```
import spacy
nlp = spacy.load('lv_core_news_sm') # Latvijas modelis
doc = nlp('Primārie: EE, LT, PL')
countries = [ent.text for ent in doc.ents if ent.label_ == 'GPE']
# Output: ['EE', 'LT', 'PL']
Confidence: 0.85
```

3. solis: Semantiskā līdzība (*Embeddings + Fuzzy Matching*)

Aprēķināt semantisko līdzību, izmantojot *sentence-transformers*:

```
from sentence_transformers import SentenceTransformer, util
model = SentenceTransformer('paraphrase-multilingual-mpnet-base-v2')

user_input_emb = model.encode('Tech-savvy millennials')
atlas_descriptions = [
    'Young professionals interested in technology',
    'Digital natives 25-35',
    'Senior citizens'
]
atlas_emb = model.encode(atlas_descriptions)
similarities = util.cos_sim(user_input_emb, atlas_emb)
# Best match: 'Digital natives 25-35' (similarity: 0.78)
Confidence: 0.70 if similarity > 0.75 else 0.50
```

4. solis: LLM rezerves mehānisms (*LLM Fallback*)

Aktivizēt, ja iepriekšējie posmi nenodrošina pārliecinošu rezultātu. Izmantot LLM ar strukturētu izvadu (JSON formāts), lai iegūtu pilnu *DimAudience* lauku komplektu.

```
PROMPT = '''
Extract audience attributes from: '{user_input}'

Return JSON:
{
  'gender': 'M/F/All',
  'age_min': int,
  'age_max': int,
  'regions': [list],
  'interests': [list],
  'reasoning': 'explanation'
}
'''

response = llm_api.complete(prompt, max_tokens=500)
Confidence: 0.60 #obligāts cilvēka apstiprinājums
```

5.2. Cilvēka līdzdalības saskarne (*Human-in-the-Loop UI*)

Ja ticamības līmenis (*confidence*) ir mazāks par 0,80, rezultāts jāuzrāda lietotājam apstiprināšanai.

Auditorijas kartēšana
nepieciešams apstiprinājums

Ievads:
Tech-savvy millennials in Baltics

AI ieteikums (68%):

- ✓ Dzimums: All
- ✓ Vecums: 25-35
- ✓ Reģioni: LV, EE, LT
- ⚠ Intereses: Technology, Innovation (nav pārliecība)

Apstiprināt

Labot

[🔍 Meklēt auditoriju manuāli](#)

5.3. Veiktspējas rādītāji un precizitāte

- Noteikumu balstīts parsētājs (Rule-Based Parser):
 - ≥70 % pārklājums vienkāršiem gadījumiem
- NLP (kombinēts):
 - ≥85 % precizitāte (*precision*) pirms cilvēka līdzdalības (HIL) validācijas
- Ar HIL validāciju:
 - 100 % precizitāte gala rezultātos

- Atbildes laiks:
 - < 3 sekundes pie viena pieprasījuma
- Ticamības līmeņa kalibrācija:
 - false positive rate < 5 %

6. Optimizācijas dzinējs

Mērķis: atrisināt daudzmērķu optimizācijas problēmu, **maksimizējot KPI svērto summu**, ievērojot **budžeta, garantiju un peļņas** ierobežojumus.

6.1. Matemātiskais formulējums

Mērķfunkcija

Maksimizēt:

$$Z = \sum (w_i \times KPI_i - penalty_guarantee - penalty_margin)$$

Kur:

- w_i = lietotāja noteikti svāri, ($\sum w_i = 1$)
- $KPI_i \in \{Reach, Awareness, Traffic, Conversions, Engagement\}$
- $penalty_guarantee$ = soda funkcija par garantiju neievērošanu
- $penalty_margin$ = soda funkcija par peļņas mērķa nesasniedzšanu

Ierobežojumi:

Budžets:	$\sum_i spend_i = budget$
Kanālu robežas:	$min\ spend_i \leq spend_i \leq max\ spend_i \quad \forall i$
Garantētie apjomi (līguma kanāliem):	$spend_i \geq guarantee_i$
Peļņas mērķis:	$margin = fee\% \cdot budget + \sum_i rebate_i - \sum_i netcost_i$ $margin \geq target\ margin$
Prime-time minimums (TV kampaņām):	$\sum prime\ spend_i \geq 0,3 \cdot budget$
Diversifikācija (riska mazināšana):	$max_i(spend_i) \leq 0,5 \cdot budget$
Ģeogrāfiskais pārklājums	$reach_j \geq min\ reach_j \quad \forall j \in regions$

6.2. Reakcijas līkņu (Response Curve) modelis

Katra KPI atkarība no izdevumiem ir nelineāra (dilstoša robežatdeve):

$$KPI_i = a_i \cdot \log(1 + b_i \cdot spend_i) + c_i + seasonality(t)$$

Kur:

- a_i, b_i, c_i = parametri, kas trenēti no vēsturiskajiem datiem (pa kanālu)
- $seasonality(t)$ = ARIMA vai Fourier komponente

Treņiņa process

1. Datu sagatavošana

- Ekstrahēt vēsturiskos **spend-KPI** pārus no **FactSpend** + **FactDelivery**
- Kopsavilkums (*aggregate*) pa nedēļām / mēnešiem
- Filtrēt atbirumus (*outliers*, $z\text{-score} > 3$)

2. Līknes pielāgošana (*Curve fitting*):

```
from scipy.optimize import curve_fit

def response_func(spend, a, b, c):
    return a * np.log(1 + b * spend) + c

params, _ = curve_fit(response_func, spend_data, kpi_data)
```

3. Sezonālitate

```
from statsmodels.tsa.arima.model import ARIMA

model = ARIMA(residuals, order=(1,0,1))
fitted_model = model.fit()
seasonal_component = fitted_model.fittedvalues
```

4. Validācija

- Train/test sadalījums: **80/20**
- **MAPE** < 15 % (test komplektā)
- $R^2 > 0.80$

6.3. Solvera (*cvxpy*) ieviešana

```
import cvxpy as cp
import numpy as np

# Lēmummainīgie
n_channels = len(channels)
spend = cp.Variable(n_channels, nonneg=True)

# KPI prognoze ar iepriekš fitētiem parametriem
kpi_predicted = []
for i, ch in enumerate(channels):
    a, b, c = params[ch]
    kpi_predicted.append(a * cp.log(1 + b * spend[i]) + c)

kpi_vec = cp.hstack(kpi_predicted) # vienā vektorā

# Maksimizēt svērto KPI summu
objective = cp.Maximize(weights @ kpi_vec)

# Ierobežojumi
constraints = [
    cp.sum(spend) == budget, # Budžets (apsver: <=)
    spend >= min_spend,     # Min uz kanālu
    spend <= max_spend,     # Max uz kanālu
```



```

    spend >= guarantees,          # Garantētie apjomi
    cp.max(spend) <= 0.5 * budget # Diversifikācija
    # prime_spend un margin var pievienot kā atvasinātas izteiksmes
]

problem = cp.Problem(objective, constraints)
problem.solve(solver=cp.ECOS, verbose=False)

if problem.status == 'optimal':
    optimal_spend = spend.value
elif problem.status == 'infeasible':
    raise ValueError('Nav realizējams risinājums – samazināt ierobežojumus')

```

DCP piezīme: $\log(\cdot)$ ir ieliekta funkcija; **Maximize** ar ieliektu mērķfunkciju ir atļauts (konkāva mērķfunkcija).

6.4. Atkāpšanās (*fallback*) mehānisms

Ja optimizators neatrod realizējamu risinājumu:

1. Pakāpeniska ierobežojumu atslābināšana:
 - Diversifikācija: **0,5 → 0,6 → 0,7**
 - Prime-time minimums: **0,3 → 0,2 → 0,1**
 - Maržas mērķis: **-10 %**, pēc tam **-20 %**
2. Heiristiska pieeja (*fallback*)
 - Budžetu sadalīt proporcionāli vēsturiskajai efektivitātei:

$$spend_i = budget \cdot \frac{historic_kpi_per_euro_i}{\sum_k historic_kpi_per_euro_k}$$

3. Cilvēka eskalācija:
 - Ja arī heiristika nedarbojas, eskalēt pie plānotāja
 - Uzrādīt konkrētu konfliktu:
 - TV3 min_spend (10k) + LTV1 min_spend (8k) > budget (15k)

7. AI plānošana viena medija letvaros

Mērķis: Optimizēt budžeta sadalījumu viena medija (TV, Radio, Web) iekšienē starp kanāliem, programmām un laika joslām.

7.1. Procesa soļi

1. **Filtrēt AdEx datus**

```

SELECT Channel, Program, Time_band, Prime_flag, Date,
       Cost_EURO, GRP, Discount_INT, NET_INT
FROM FactSpend
WHERE Medium = 'TV' AND Date >= '2024-01-01'

```

2. Grupēt un aprēķināt efektivitāti

```
SELECT Channel, Program, Time_band,
       AVG(GRP / NULLIF(Cost_EURO, 0)) AS grp_per_euro,
       AVG(Discount_INT) AS avg_discount,
       SUM(CASE WHEN Prime_flag=1 THEN 1 ELSE 0 END) / COUNT(*) AS
prime_ratio
FROM filtered_data
GROUP BY Channel, Program, Time_band
```

3. Reakcijas līkņu pielāgošana (*Fit Response Curves*)

- Katram kanālam: $KPI_i = a_i \cdot \log(1 + b_i \cdot spend_i) + c_i$
- Sezonālitate: tiek piemēroti mēneša un nedēļas dienas koeficienti, kas korelē ar auditorijas uzvedības svārstībām.

4. Optimizācija ar ierobežojumiem

Izmantojot iepriekš modelētās reakcijas līknes un definētos komerciālos noteikumus (budžets, garantijas, *prime-time* u.c.), tiek atrasts optimālais izdevumu sadalījums pa kanāliem un programmām, maksimizējot KPI.

Maximize: $\sum (grp_per_euro_i \times spend_i \times (1 + discount_i))$

Subject to:

$\sum spend_i = budget_tv$
 $spend_i \geq min_spend_i$ (garantijas)
 $\sum prime_spend_i \geq 0.3 \times budget_tv$ (prime-time minimums)
 $discount_i \in valid_discount_bands$ (pēc rate card)

5. Rezultātu ģenerācija

- **Sadalījums pa kanāliem:** budžeta, GRP un reach prognozes katram kanālam
- **Programmu / laika joslu rekomendācijas:** optimālais īpatsvars pēc efektivitātes
- **Izskaidrojums (explainability):** paskaidrojums, kāpēc izvēlēts konkrētais sadalījums — aktīvie ierobežojumi, efektivitātes faktori un ietekmes svāri

7.2. Izvada piemērs

```
{
  "medium": "TV",
  "budget_eur": 20000,
  "split": [
    {
      "channel": "TV3",
      "spend": 8500,
      "expected_grp": 96.4,
      "programs": ["Ziņas", "Prime series"],
      "time_bands": ["19:00-22:00 (55%)", "22:00-00:00 (30%)"]
    },
    {
      "channel": "LTV1",
      "spend": 6500,
      "expected_grp": 78.1
    },
    {
      "channel": "STV",
```

```
    "spend": 3000,  
    "expected_grp": 35.9  
  },  
  {  
    "channel": "Helio bundle",  
    "spend": 2000,  
    "expected_grp": 22.7  
  }  
],  
"expected_total_grp": 233.1,  
"expected_reach": 0.58,  
"explain": [  
  "TV3 offers best GRP/€ ratio (11.3) in target segment",  
  "LTV1 chosen for reach diversification (+12% unique reach)",  
  "Discount band >8% unlocked for Helio group (€8.5k spend)",  
  "Prime-time allocation: 62% (above 30% minimum)"  
]  
}
```

8. Izskaidrojamība (*explainability*)

Mērķis:

Nodrošināt pilnīgu pārredzamību par to, **kāpēc AI izvēlējās konkrēto budžeta sadalījumu**, un dot iespēju cilvēkam saprast, pārbaudīt un auditēt optimizācijas rezultātu.

8.1. Izskaidrojuma komponentes

1. Aktīvie ierobežojumi (*Binding Constraints*)

- Identificēt, kuri ierobežojumi bija *aktīvi* (sasniedza savu limitu).
- Piemēri:
 - “Budžets pilnībā izmantots (€50,000)”
 - “TV3 minimālā garantija sasniegta (€10k)”

2. Faktoru ietekme (*Feature Importance*)

Nosaka, kuri mainīgie visvairāk ietekmēja gala sadalījumu.

Galvenie ietekmētāji (*Top Drivers*):

- GRP/€ efektivitāte - 42 %
- Atlaides aktivizēšana - 28 %
- Reach diversifikācija - 18 %
- Garantētie apjomi - 12 %

3. Pretfaktuālā analīze (*Counterfactual Analysis*)

Analizē hipotētiskus “kas notiktu, ja...” scenārijus, lai parādītu modeļa jutīgumu pret izmaiņām.

Piemērs:

“Ja budžets palielinātos par 10 %, sagaidāmais GRP pieaugtu par 8.3, galvenokārt caur LTV1.”

4. LLM stāstījums (*LLM Narrative*)

Automātiski ģenerēts skaidrojums dabīgā valodā, izmantojot GPT-5 vai Claude.

Mērķis — radīt cilvēkam saprotamu kopsavilkumu un pamatot izvēlēto sadalījumu.

8.2. LLM stāstījuma *pipeline*

EXPLAIN_PROMPT = '''

You are an expert media planner. Explain this optimization result:

Budget: €50,000

KPI Target: Reach 60% + Awareness 40%

Optimal Split:

- TV: €20,000 (40%) → Expected GRP: 233
- Radio: €12,000 (24%) → Expected GRP: 145
- Web: €18,000 (36%) → Expected Reach: 42%

Binding Constraints:

- Budget fully utilized
- TV min guarantee (€10k) active

Feature Importance:

1. GRP/€ efficiency (42%)

2. Reach diversification (28%)
3. Discount unlocking (18%)

Provide:

1. Executive summary (2-3 sentences)
2. Key rationale (3-5 bullet points)
3. Potential risks/limitations

8.3. LLM stāstījuma izvada piemērs

Kopsavilkums (*Executive Summary*):

Optimizācija piešķir 40 % TV plašas atpazīstamības nodrošināšanai, 36 % tīmekļa kanāliem mērķētam reach, un 24 % radio frekvences palielināšanai. Šāds sadalījums maksimizē kopējo sasniedzamību un atpazīstamību, vienlaikus saglabājot TV garantijas un uzlabojot GRP/€ efektivitāti par 11 % salīdzinājumā ar vēsturiskajiem datiem.

Galvenie pamatojumi (*Key Rationale*):

- TV (€20k) nodrošina vislielāko GRP (233) atpazīstamības veidošanai; TV3 kanāls piedāvā vislabāko efektivitāti (11.3 GRP/€)
- Web (€18k) nodrošina +42 % papildu reach, īpaši 25–40 gadu vecuma segmentā
- Radio (€12k) sniedz izmaksu ziņā efektīvu frekvences pieaugumu (145 GRP) pie €82/GRP (salīdzinājumā ar TV €86/GRP)
- “Helio bundle” 8.5 % atlaide aktivizēta, sasniedzot €8.5k kopējo apjomu visos kanālos

Riski un ierobežojumi (*Risks & Limitations*):

- Web sasniedzamība balstās uz trešo pušu sīkdatņu datiem, kuru pieejamība var samazināties 2025. gadā
- TV garantija fiksē €100k neatkarīgi no sezonālajām efektivitātes izmaiņām

9. API Specifikācija

RESTful API ar FastAPI, OAuth2 autentifikāciju un automātisku OpenAPI dokumentāciju.

9.1. Pamatendpunkti (*Core Endpoints*)

POST /api/v1/campaigns/optimize

Ģenerē optimālu mediju sadalījumu.

Pieprasījums:

```
{
  "campaign_id": "string",
  "budget": 50000,
  "kpi_targets": {
    "reach": {
      "weight": 0.6,
      "target": 0.65
    },
    "awareness": {
      "weight": 0.4,
      "target": null
    }
  },
  "audience": "Sievietes 25-45, Rīga un Pierīga",
  "constraints": {
    "min_spend_tv": 10000,
    "max_spend_web": 20000,
    "margin_target_pct": 12
  },
  "start_date": "2025-11-01",
  "end_date": "2025-12-31"
}
```

Atbilde (200 OK):

```
{
  "plan_id": "uuid",
  "status": "optimal",
  "split": [
    {
      "medium": "TV",
      "spend": 20000,
      "expected_kpi": {}
    },
    {
      "medium": "Radio",
      "spend": 12000,
      "expected_kpi": {}
    },
    {
      "medium": "Web",
      "spend": 18000,
      "expected_kpi": {}
    }
  ],
  "total_expected_reach": 0.58,
}
```

```

"total_expected_grp": 378,
"explainability": {
  "executive_summary": "...",
  "key_rationale": [],
  "binding_constraints": []
},
"solve_time_ms": 2847,
"created_at": "2025-10-10T14:23:15Z"
}

```

POST /api/v1/audiences/map

Kartē auditorijas `string` uz `DimAudience`.

Pieprasījums:

```

{
  "audience_string": "Tech-savvy millennials in Baltics"
}

```

Atbilde (200 OK):

```

{
  "audience_id": "ATL-BAL-A-25-35-TECH",
  "confidence": 0.68,
  "attributes": {
    "gender": "All",
    "age_min": 25,
    "age_max": 35,
    "regions": ["LV", "EE", "LT"],
    "interests": ["Technology", "Innovation"]
  },
  "alternatives": [
    {
      "audience_id": "...",
      "confidence": 0.54
    },
    {
      "audience_id": "...",
      "confidence": 0.49
    }
  ],
  "requires_review": true
}

```

GET /api/v1/campaigns/{id}/performance

Salīdzina plānoto mediju apjomu pret piegādāto.

Atbilde (200 OK):

```

{
  "campaign_id": "uuid",
  "comparison": [
    {
      "medium": "TV",
      "planned_spend": 20000,
      "actual_spend": 19500,
      "planned_grp": 233,

```

```

    "actual_grp": 228,
    "variance_pct": -2.1
  }
],
"overall_mape": 3.7,
"status": "within_tolerance"
}

```

9.2. Autentifikācija

Nodrošināt drošu un kontrolētu piekļuvi API resursiem, izmantojot standarta autentifikācijas un autorizācijas mehānismus.

- **OAuth2 ar Bearer tokeniem**

Galvenais autentifikācijas veids. Katram pieprasījumam jāiekļauj HTTP galvene:

`Authorization: Bearer <access_token>`

- **API atslēgas (API keys)**

Sekundārs piekļuves mehānisms, paredzēts sistēmu integrācijām vai iekšējiem servisiem.

Lietotāju piekļuves kontrole

- Lomu sadalījums (*Role-based access*):
 - **Admin** – pilna piekļuve, konfigurācijas un lietotāju pārvaldība
 - **Planner** – piekļuve plānošanas, optimizācijas un auditoriju datiem
 - **Viewer** – tikai lasīšanas piekļuve (plānu un KPI skatīšana)

Drošības ierobežojumi

- Pieprasījumu ierobežošana (Rate limiting):
 - Maksimāli 100 pieprasījumi minūtē uz vienu lietotāju.
- Pārsniedzot limitu, API atgriež HTTP statusu `429 Too Many Requests`.

9.3. Kļūdu atbildes (*Error Responses*)

Mērķis:

Definēt vienotu kļūdu atbilžu struktūru, lai klienti varētu konsekventi apstrādāt API kļūdas un lietotāji saņemtu saprotamus paskaidrojumus un ieteikumus.

Formāts

Visas kļūdu atbildes tiek atgrieztas **JSON formātā** ar šādiem laukiem:

- `error` – īss kļūdas tips vai kods
- `details` – papildu skaidrojums par problēmu
- `suggestion` (*neobligāts*) – ieteikums, kā to novērst

Piemēri

400 Bad Request

Pieprasījums ir sintaktiski pareizs, bet satur nederīgus datus.


```
{  
  "error": "Invalid budget",  
  "details": "Budget must be positive"  
}
```

422 Unprocessable Entity

Pieprasījums ir loģiski nekorekts — optimizāciju nav iespējams izpildīt ar norādītajiem parametriem.

```
{  
  "error": "Infeasible optimization",  
  "details": "Guarantees exceed budget",  
  "suggestion": "Reduce min_spend_tv or increase budget"  
}
```

401 Unauthorized

Nav derīga autentifikācijas pilnvara

```
{  
  "error": "Unauthorized",  
  "details": "Missing or invalid token"  
}
```

403 Forbidden

Lietotājam nav tiesību pieprasīt šo resursu

```
{  
  "error": "Access denied",  
  "details": "Role Viewer cannot modify plans"  
}
```

500 Internal Server Error

Iekšēja sistēmas kļūda

```
{  
  "error": "Server error",  
  "details": "Unexpected exception in optimizer"  
}
```

10. Monitorings un veikspēja

10.1. Galvenie rādītāji

Rādītājs	Mērķis	Brīdinājums
Optimizācijas izpildes laiks	p95 < 30s	p95 > 45s
Modeļa precizitāte (MAPE)	< 15%	> 20%
API atbildes laiks	p99 < 5s	p99 > 10s
ETL datu aktualitāte	< 2 stundas	> 4 stundas
Optimizācijas veiksmes rādītājs	> 95%	< 90% (24h)
Budžeta izmantošanas rādītājs	95-105%	< 90% vai > 110%
Peļņas sasniegšanas rādītājs	> mērķa -2%	< mērķa -5%

10.2. Informācijas paneļa komponentes

- Reāllaika sistēmas stāvoklis:** ETL statuss, API pieejamība, optimizācijas rinda
- Modeļa veikspēja:** MAPE dinamika, precizitāte pa mediju tiptiem, nobīdes (drift) detekcija
- Biznesa rādītāji:** ģenerēto plānu skaits, vidējais optimizācijas laiks, veiksmes īpatsvars
- Anomāliju detekcija:** neparasti tēriņu modeļi, KPI novirzes, ierobežojumu pārkāpumi

10.3. Brīdinājumu stratēģija

Kritiski (SMS + Tālrunis)

- API nepieejamība > 1 min
- ETL kļūme > 2 secīgi izpildes cikli
- Datu bāzes savienojuma zudums

Augsts (SMS)

- Optimizācijas veiksmes rādītājs < 90% (24 h)
- MAPE > 20% konkrētā kampaņu tipā
- ETL aizkave > 4 stundas

Vidēja prioritāte (e-pasts)

- Modeļa nobīde (precizitātes samazinājums > 5% nedēļā)
- Neparasts ierobežojumu modelis (>10 neizpildāmi gadījumi 1 h laikā)

11. Testēšana un validācija

11.1. Vienību testi (*Unit tests*)

- **NLP kartējumi:** 100+ testgadījumi ar zināmiem ievades/izvades datiem
- **Atlaides funkcijas:** robežgadījumi (0, maksimums, robežvērtības)
- **Maržas aprēķini:** dažādas komisijas struktūras (% , fiksēta, hibrida)
- **Reakcijas līknes:** dilstošās atdeves validācija

11.2. Integrācijas testi (*Integration tests*)

- **ETL cauruļvads:** pilns cikls AdEx → DW → validācija
- **API darbplūsmas:** POST /optimize → GET /performance ķēde
- **Optimizators un ierobežojumi:** pārbaudīt, vai visi ierobežojumu tipi tiek ievēroti

11.3. Atpakaļtestēšana (*Backtesting*)

AI plānu salīdzināšana ar vēsturiskajiem cilvēka veidotajiem plāniem:

1. Atlasīt 20 kampaņas
2. Nodrošināt dažādus reklāmdevējus, budžetus un KPI kombinācijas
3. Nodrošināt pārklājumu 2024. gada Q1–Q4 periodā
4. Pārpalāst optimizāciju, izmantojot tikai datus līdz kampaņas sākuma datumam
5. Ģenerēt AI plānu ar tādiem pašiem ierobežojumiem
6. Salīdzināt rezultātus

Vērtēšanas rādītāji:

- KPI sasniegums: AI vs cilvēka plāni
- Izmaksu efektivitāte: GRP/€ salīdzinājums
- Ierobežojumu ievērošana: pārkāpumu skaits

Veiksmes kritēriji:

- $\geq 60\%$ kampaņu: AI rezultāts $\pm 10\text{--}15\%$ robežās no faktiskā
- $\geq 10\%$ labāka GRP/€ efektivitāte salīdzinājumā ar cilvēka plāniem

11.4. Zelta datu kopas (*Golden Datasets*)

Trīs etalonkampaņas reproducējamībai:

1. **Vienkārša TV kampaņa**
 - Budžets: €15 000
 - Mērķauditorija: Sasniedzamība 50%, sievietes 25–45
 - Optimālais sadalījums: TV3 60%, LTV1 40%
2. **Daudzkanālu kampaņa ar garantijām**
 - Budžets: €50 000
 - KPI: Sasniedzamība 60%, atpazīstamība 40%

- Ierobežojumi: TV min. €10 000, radio max. €15 000

3. Kompleksa kampaņa ar maržas optimizāciju

- Budžets: €100 000
- Vairāki KPI, mērķa marža 12%
- Atlaides diapazoni, atlaides (rebates), diversifikācijas nosacījumi

12. Ieviešana, tai skaitā pakāpeniska ieviešana

12.1. Pakāpeniskas ieviešanas plāns

1. fāze – *Shadow mode* (3 nedēļas)

- AI ģenerē plānus paralēli cilvēka veidotajiem plāniem
- Neviens AI plāns netiek izmantots produkcijas vidē
- Tiek veikta AI un cilvēka prognožu salīdzināšana
- **Mērķis:** identificēt būtiskas novirzes vai kļūdas (major *discrepancies*)

2. fāze – Pilotprojekts (2 mēneši)

- 20% kampaņu tiek veidotas, izmantojot AI plānus (ar cilvēka pārbaudi)
- Atlasītās kampaņas: zems risks, mazs budžets (< €20 000)
- Ikdienu monitorings un iknedēļas pārskati
- Ieviests atgriezeniskās saites cikls (feedback loop): plānotāji var labot vai noraidīt AI ieteikumus

3. fāze – Paplašināšana (2 mēneši)

- 50% kampaņu tiek plānotas ar AI palīdzību
- Iekļautas vidēja budžeta kampaņas (€20 000–€50 000)
- Samazināta uzraudzība – notiek tikai izlases pārbaudes (*spot checks*)

4. fāze – Pilnīga automatizācija (3+ mēneši)

- 80% kampaņu tiek pilnībā automatizētas
- AI plāni ar augstu pārliecības līmeni (>0.85) tiek automātiski apstiprināti
- Cilvēka pārbaude tiek veikta tikai:
 - augsta budžeta kampaņām (> €100 000),
 - zemas pārliecības (*low-confidence*) gadījumiem,
 - anomālijām (neparasti rezultāti vai ierobežojumu pārkāpumi)

12.2. Atsaukšanas (rollback) nosacījumi

Sistēma tiek atgriezta iepriekšējā (drošā) stāvoklī, ja tiek konstatēts viens vai vairāki no šiem rādītājiem:

- MAPE > 25% trīs secīgas dienas
- Optimizācijas veiksmes rādītājs < 85%
- Vairāk nekā 5 plānotāju manuāli labojumi dienā (indikācija par sistēmisku problēmu)
- Klienta eskalācija gadījumā, ja budžeta pārsniegums > 10%

13. Riska faktori un mazināšanas pasākumi

Risks	Ietekme	Mazinošie pasākumi
Kantar Atlas datu nepietiekama granularitāte	Auditoriju kartēšanas precizitāte < 85%	Obligāta HIL (<i>Human-in-the-Loop</i>) validācija gadījumos ar zemu pārliecības līmeni; iespēja papildināt ar trešo pušu datiem (piem., Gemius)
Atlaižu sarežģītība (<i>tier</i> struktūra)	Optimizators var nepareizi aprēķināt neto izmaksas	Detalizēta SCD2 versēšana tabulā <i>FactCommercials</i> ; vienību testi visām atlaides līknēm un diapazoniem
<i>Traffic / Conversion</i> KPI bez 1st-party datiem	Reakcijas līknes kļūst neuzticamas digitālajiem KPI	v1.0: fokuss uz Reach / Awareness / GRP KPI; v1.1: integrēt GA4 un Meta Pixel datus
Modeļa nobīde (model drift) tirgus izmaiņu dēļ	Prognožu precizitāte samazinās laika gaitā	Mēneša modeļa atkārtota apmācība, nobīdes (drift) brīdinājumi, ceturkšņa atpakaļtesti
Neizpildāmi optimizācijas scenāriji (infeasible)	Ierobežojumi nav izpildāmi → plāns netiek ģenerēts	3 līmeņu rezerves mehānisms: ierobežojumu atslābināšana → heuristiskā pieeja → cilvēka eskalācija ar automātisku kļūmes aprakstu
Iesaistīto pušu pretestība (stakeholder resistance)	Plānotāji neuzticas AI ieteikumiem	Skaidrojamu rezultātu sistēma (<i>Explainability</i>), pakāpeniska ieviešana, apmācību sesijas, agrīnu uzvaru demonstrēšana (efektivitātes pieaugums)

14. Panākumu kritēriji

14.1. Tehniskie kritēriji (MVP)

Kritērijs	Mērķis
ETL precizitāte	≥ 98% ierakstu bez kļūdām
NLP kartēšanas precizitāte (ar HIL)	100% pēc cilvēka apstiprināšanas
Optimizācijas izpildes laiks	≤ 30 sekundes (p95)
Budžeta un ierobežojumu ievērošana	100% (visi ierobežojumi izpildīti)
Vienas medija plāna ģenerēšanas laiks	< 30 sekundes ar izskaidrojumu (explainability)

14.2. Biznesa kritēriji

Kritērijs	Mērķis
Atpakaļtestēšanas precizitāte (MAPE)	≥ 60% kampaņu ±10–15% robežās no faktiskajiem rezultātiem
Efektivitātes pieaugums salīdzinājumā ar manuālo plānošanu	≥ 10% labāka GRP/€ efektivitāte
Plānošanas laika samazinājums	50% ietaupījums (no 2 h uz 1 h uz vienu plānu)
Ieinteresēto pušu (plānotāju) apmierinātība	Vidējais vērtējums ≥ 4/5 pēc aptaujas
Sasniegtā peļņas norma	≥ 90% kampaņu sasniedz peļņas mērķi ±2% robežās

15. Nākotnes uzlabojumi

15.1. Versija v1.1 (plānots 2026. gada 3. ceturksnis)

- Platformu līmeņa plānošana: integrācija ar Facebook Ads, Google Ads un TikTok API
- RTB validācija: reāllaika solījumu (*real-time bid*) pārbaude
- Datu integrācija ar Google 360, GA4 un Meta lai uzlabotu konversiju reakcijas līknes (*conversion response curves*)
- Radošuma efektivitāte: reklāmas materiālu (*ad creatives*) novērtēšanas un ieteikumu sistēma

15.2. Versija v1.2 (plānots 2026. gada 4. ceturksnis)

- MMM-lite modelis: *Marketing Mix Modeling* vieglā versija, lai noteiktu pamata efektus (*baseline effects*)
- Incrementalitātes testēšana: *holdout* testi un reklāmas pacēluma (*lift*) mērīšana
- Budžeta tempu kontrole (budget pacing): kampaņas ietvaros veiktas dienas/nedēļas budžeta korekcijas
- Daudzkanālu atribūcija (*multi-touch attribution*): starpkanālu lietotāju ceļa (journey) modelēšana

15.3. Versija v2.0 (plānots 2027. gada 1. ceturksnis)

- Stiprinošā mācīšanās (reinforcement learning): AI modelis mācās no kampaņu rezultātiem un automātiski uzlabo plānošanas stratēģiju
- Scenāriju plānošana: "What-if" analīzes saskarne (UI) budžeta un tirgus izmaiņu simulācijai
- Konkurences izlūkošana: konkurentu tēriņu uzraudzība un pret-stratēģiju (counter-strategy) ģenerēšana
- Pilna programmatiskā integrācija: DSP API integrācija automatizētai reklāmas iegādei

16. Pielikumi

16.1. String → Kantar Atlas kartēšanas piemēri

User Input	Interpretētie (parsētie) atribūti	Ģenerētais Audience ID
All 18-43	gender=All, age=[18,43]	ATL-LV-A-18-43
Sievietes 20-50, Rīga un Pierīga	gender=F, age=[20,50], geo=[Rīga, Pierīga]	ATL-LV-F-20-50-RP
Primārie: EE, LT, PL	countries=[EE, LT, PL]	ATL-BALTIC-A-ALL

16.2. Atlaižu līknes piemērs

TV atlaižu struktūra (SCD2)

Budžeta intervāls	Atlaide %	No	Līdz
€0 - €20k	5%	2024-01-01	2025-12-31
€20k - €50k	9%	2024-01-01	2025-12-31
€50k+	12%	2024-01-01	2025-12-31

17. Terminu vārdnīca

Termins	Definīcija
GRP	<i>Gross Rating Point</i> — auditorijas reitinga punkts, kas raksturo reklāmas kopējo pārklājumu (sasniežamība × frekvence).
Reach	Auditorijas sasniežamība — procentuālais rādītājs, kas norāda, cik % mērķauditorijas redzēja reklāmu vismaz vienu reizi.
MAPE	<i>Mean Absolute Percentage Error</i> — vidējā absolūtā procentuālā kļūda starp prognozētajām un faktiskajām vērtībām.
SCD Type 2	<i>Slowly Changing Dimension Type 2</i> — datu noliktavas versēšanas metode, kas saglabā katras vērtības vēsturiskās versijas ar “Valid From / Valid To”.
HIL	<i>Human-in-the-Loop</i> — pieeja, kur cilvēks validē vai koriģē mākslīgā intelekta ģenerētos rezultātus.
Prime-time	Galvenais skatīšanās laiks televīzijā (parasti no 19:00 līdz 23:00), kad auditorijas apjoms ir vislielākais.
Response curve	Atbildes līkne — matemātisks modelis, kas parāda, kā KPI (piemēram, GRP vai sasniežamība) mainās atkarībā no reklāmas izdevumiem.
Binding constraint	Aktīvais ierobežojums — ierobežojums, kas nosaka optimālo risinājumu, jo ir “sasniegjis savu limitu” un ietekmē modeļa rezultātu.

— Dokumenta beigas —