

Buhalterijos AI Asistentas

Projekto dokumentacija

Autorius: Dominykas Juodvalkis

Data: 2025 m. gruodžio 3 d.

Turinys

1	Įvadas	2
2	Terminai ir santrumpos	2
3	Projekto apžvalga	2
3.1	Tikslai	2
3.2	Sistemos ribos	2
4	Reikalavimai	2
4.1	Funkciniai reikalavimai	2
4.2	Nefunkciniai reikalavimai	3
5	Funkcijų aprašymas	3
5.1	OCR apdorojimas	3
5.2	Teksto analizė (NLP)	3
5.3	ML klasifikacija	3
5.4	Apskaitos įrašų generavimas	3
6	Sistemos architektūra	4
6.1	UML Diagramos	4
6.1.1	Klasės diagrama	4
6.1.2	Veiklos diagrama	4
6.1.3	Sekų diagrama	4
6.1.4	Komponentų diagrama	4
7	Duomenų srautai ir mokymo ciklai	4
8	Technologijos ir įrankiai	6
9	Išvados	6

1 Įvadas

Buhalterijos AI Asistentas yra sistema, skirta finansinių dokumentų analizės, OCR ir apskaitos įrašų generavimo automatizavimui. Ji apjungia pažangias dirbtinio intelekto technologijas, kad sumažintų rankinio darbo kiekį ir padidintų duomenų tikslumą.

2 Terminai ir santrumpos

- **OCR (Optical Character Recognition)** – optinis simbolių atpažinimas.
- **NLP (Natural Language Processing)** – natūralios kalbos apdorojimas.
- **ML (Machine Learning)** – mašininis mokymasis.
- **GUI** – grafinė vartotojo sąsaja.

3 Projekto apžvalga

3.1 Tikslai

- Automatizuoti finansinių dokumentų analizę ir apdorojimą.
- Struktūrizuoti tekstinę informaciją NLP pagalba.
- Naudoti ML modelius dokumentų klasifikavimui.
- Generuoti apskaitos įrašus automatiškai.

3.2 Sistemos ribos

Sistema apdoroja PDF, skenuotas sąskaitas ir kvitus, bet nebus atsakinga už rankinius įrašus, nefinansinių dokumentų klasifikaciją ar juridinių sprendimų priėmimą.

4 Reikalavimai

4.1 Funkciniai reikalavimai

1. Sistema turi priimti skenuotus dokumentus (PDF ar nuotraukų formatu).
2. OCR modulis turi konvertuoti dokumentus į tekstą su bent 95% tikslumu.
3. NLP modulis turi struktūrizuoti tekstą į duomenų lenteles.
4. ML modulis turi klasifikuoti dokumentus pagal tipą (pvz., kvitas, sąskaita).
5. Sistema turi generuoti apskaitos įrašus pagal nustatytas taisykles.
6. Sistema turi registruoti ir saugoti apdorotus dokumentus.

4.2 Nefunkciniai reikalavimai

- Naudotojo sąsaja turi būti intuityvi ir suprantama.
- Sistema turi būti suderinama su Windows ir Linux operacinėmis sistemomis.
- Atsakymo laikas apdorojant dokumentą turi būti mažesnis nei 10 sekundžių.
- Sistema turi užtikrinti duomenų konfidencialumą ir saugumą.

5 Funkcijų aprašymas

5.1 OCR apdorojimas

- Įėjimas: PDF arba nuotrauka
- Išėjimas: tekstas
- Aprašymas: modulis naudoja OCR technologiją (pvz., Tesseract), kad iš dokumento išgautų tekstinę informaciją.

5.2 Teksto analizė (NLP)

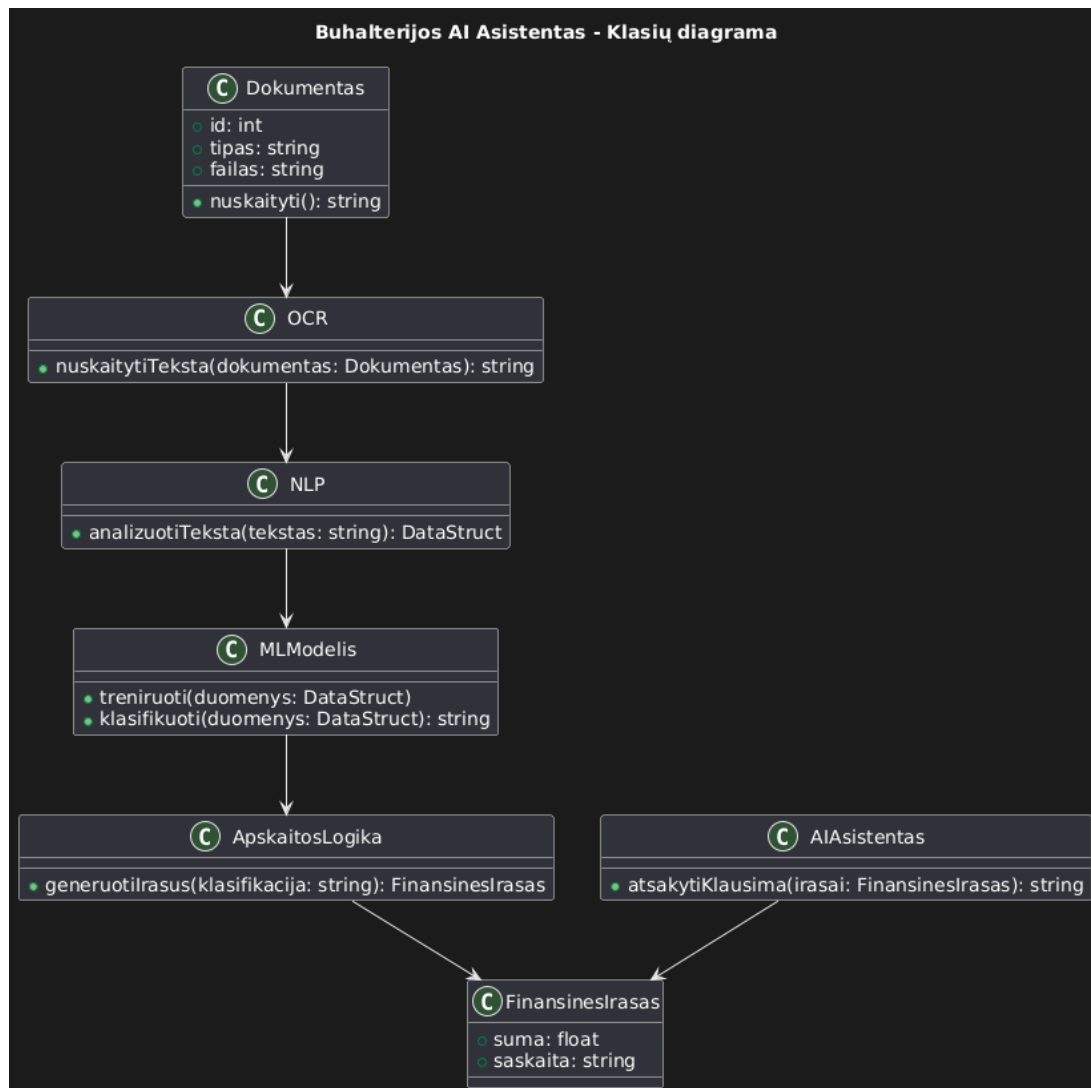
- Įėjimas: OCR rezultatas (tekstinė informacija)
- Išėjimas: struktūrizuoti duomenys (lentelės)
- Aprašymas: modulis identifikuoja reikšmingus duomenis (sumas, datas, PVM, sąskaitos numerį) ir paverčia juos struktūrizuota forma.

5.3 ML klasifikacija

- Įėjimas: struktūrizuoti duomenys
- Išėjimas: dokumento tipas
- Aprašymas: modulis klasifikuoja dokumentus, pvz., kvitas, sąskaita, banko pervedimas, naudodamas mokymosi modelius.

5.4 Apskaitos įrašų generavimas

- Įėjimas: klasifikuoti ir struktūrizuoti duomenys
- Išėjimas: apskaitos įrašai
- Aprašymas: generuoja įrašus pagal verslo taisykles, pvz., debetas, kreditas, PVM apskaičiavimas.



1 pav.: Klasės diagrama

6 Sistemos architektūra

6.1 UML Diagramos

6.1.1 Klasės diagrama

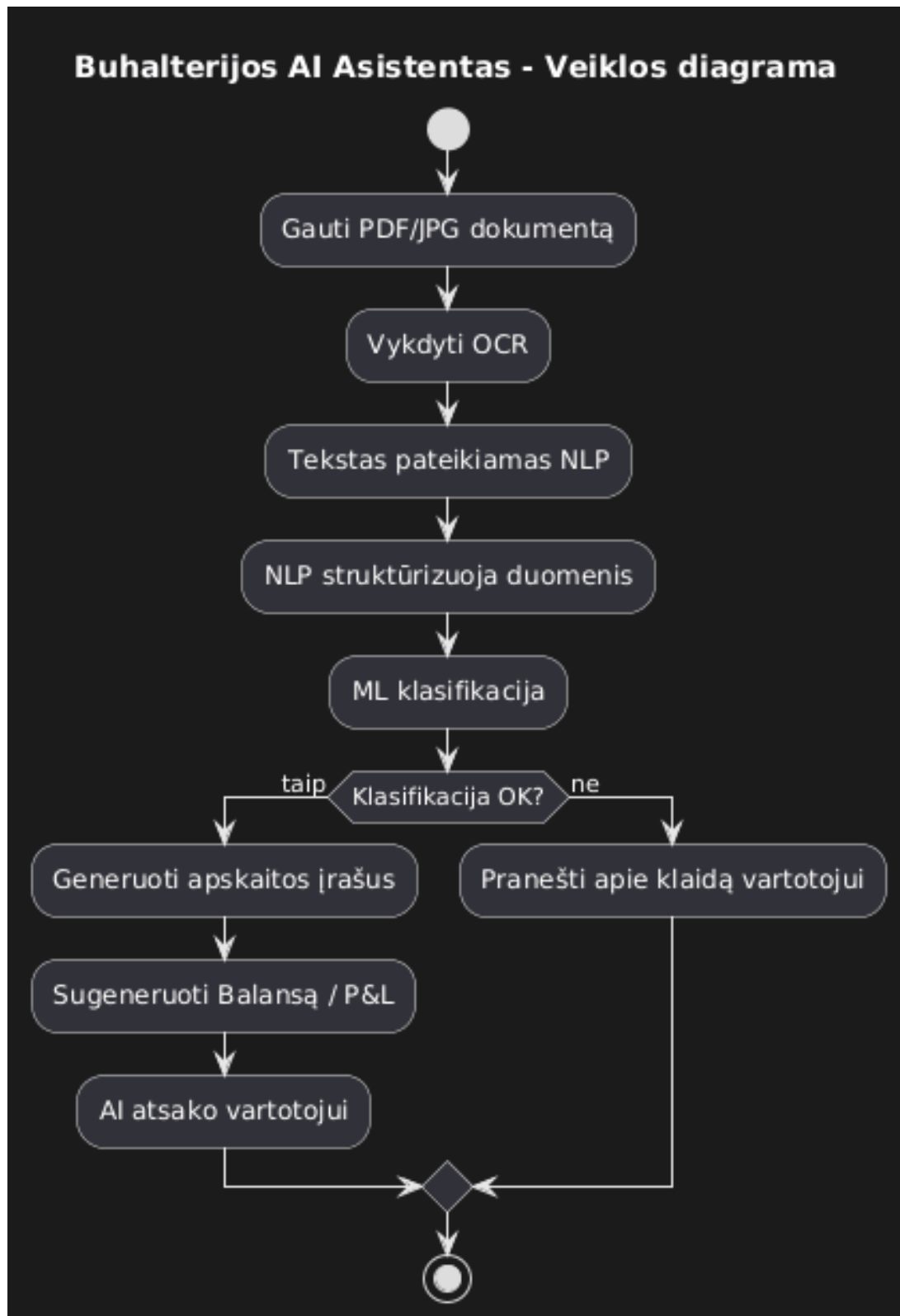
6.1.2 Veiklos diagrama

6.1.3 Sekų diagrama

6.1.4 Komponentų diagrama

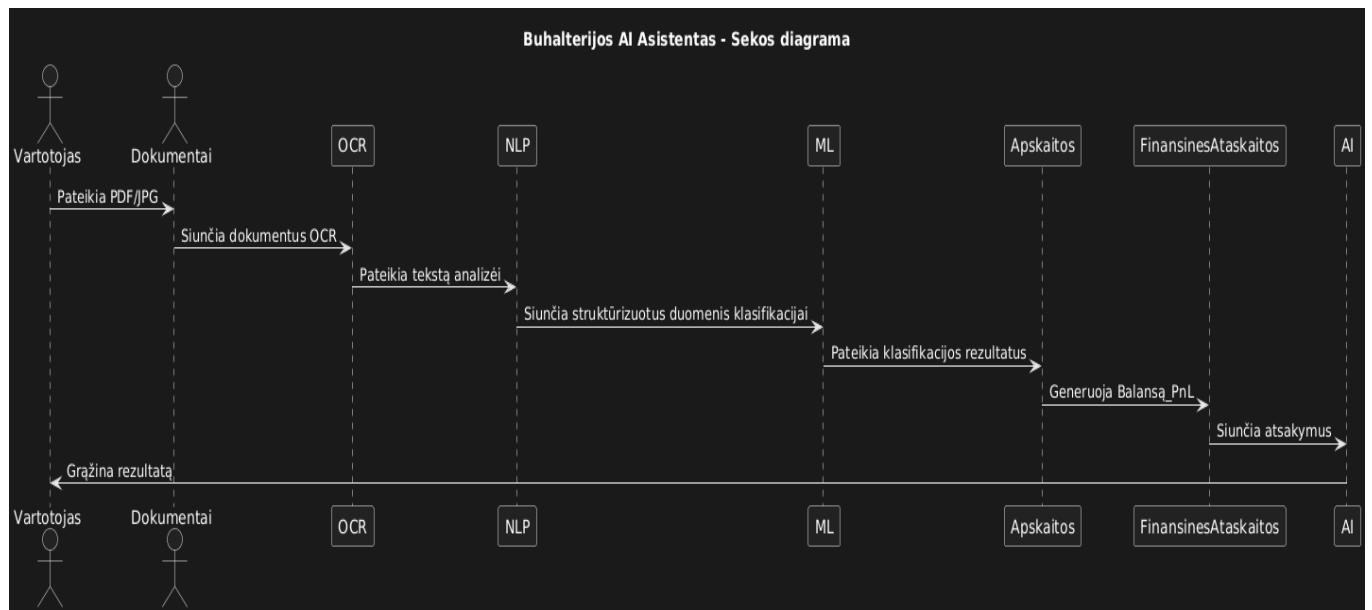
7 Duomenų srautai ir mokymo ciklai

1. Train #1 – pradinio modelio mokymas.
2. Train #2 – modelio tobulinimas.
3. Train #3 – galutinio modelio optimizavimas.



2 pav.: Veiklos diagrama

4. Test #4 – modelio testavimas su naujais duomenimis.



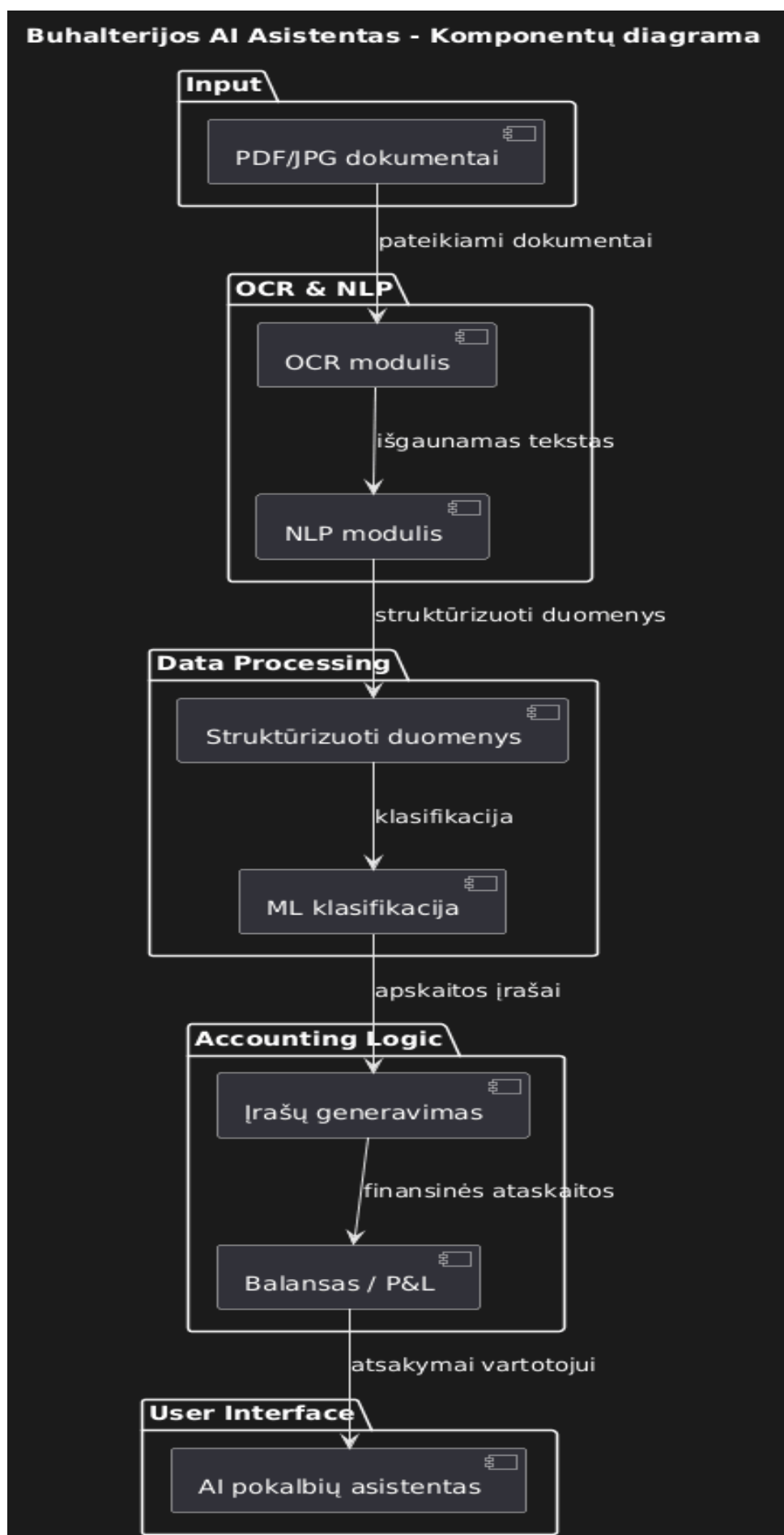
3 pav.: Sekų diagrama

8 Technologijos ir įrankiai

- Python
- Google Generative AI / Gemini
- BeautifulSoup
- OCR biblioteka (pvz., Tesseract)
- LaTeX dokumentacijos kūrimui

9 Išvados

Buhalterijos AI Asistentas leidžia automatizuoti finansinių dokumentų analizę ir apskaitos įrašų generavimą. Sistema sumažina rankinio darbo kiekį, užtikrina didesnę tikslumą ir leidžia efektyviau apdoroti finansinius duomenis.



4 pav.: Komponentų diagrama