

AI 회계시스템 소개

자동분개 · 거래매칭 · 세액공제

(주)대유씨엔에이
서울대학교 통계학과

Contents

1. AI회계 개요

2. 기대효과

3. AI 회계시스템

- 1) 자동분개(계정과목 분류)
- 2) AI매칭(내부거래 및 미결관리)
- 3) 세액공제 분류

4. 도입 시나리오

5. 향후과제

6. 참고

- 1) AI매칭알고리즘(내부거래 및 미결관리)

1. AI회계의 개요



I. 수동프로세스의 자동화

2026년 AI가 재무분야에서 사람을 대체할 것

II. Zero Day Close

사전Risk관리, 업무 프로세스 효율화

III. 내부통합 비즈니스 환경

실시간 통합 데이터 기반의 의사결정

1. AI회계의 개요

- AI회계 업무 분야는 크게 다음 6가지로 나눌 수 있다.

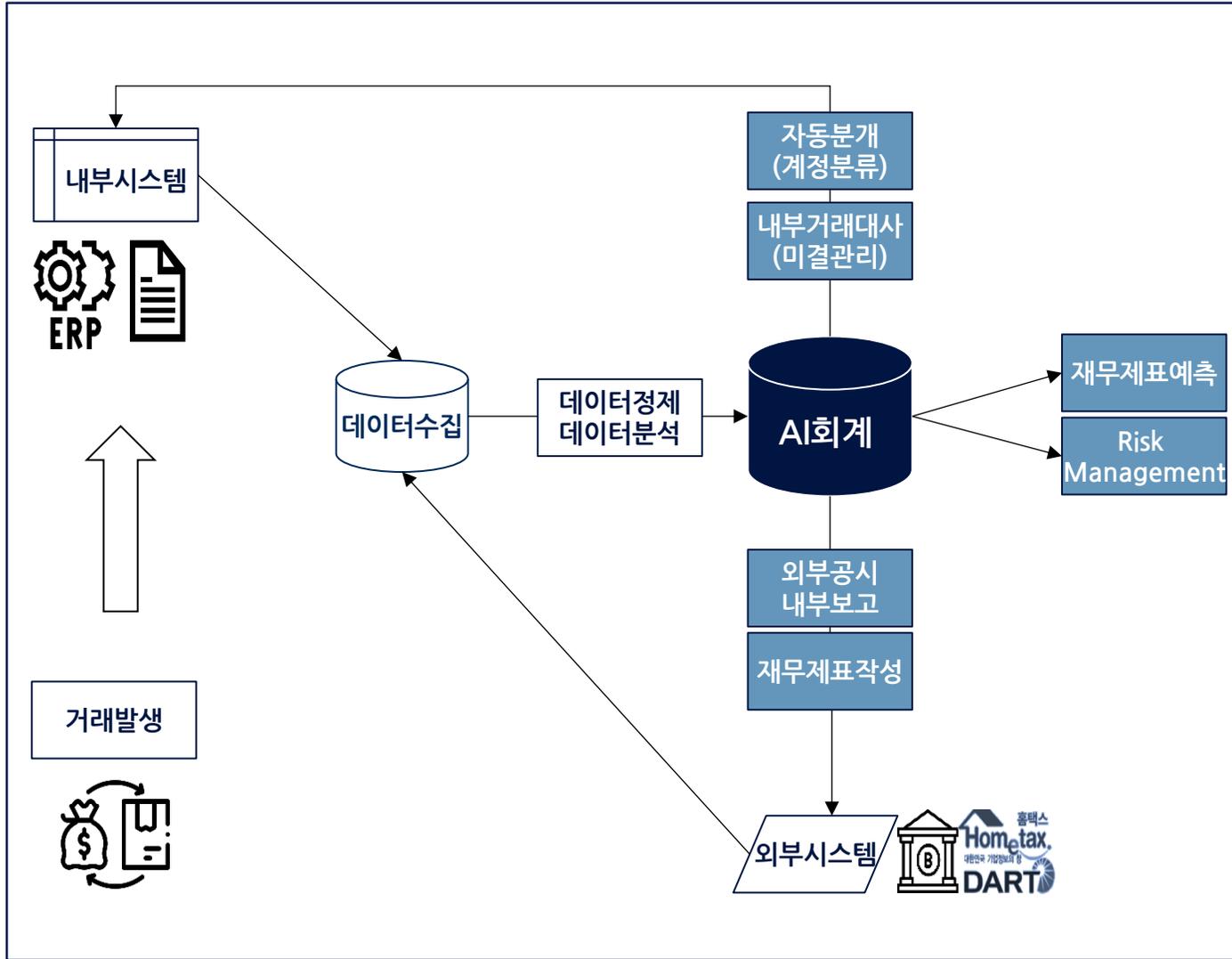
구분	업무	알고리즘	고려사항
Book keeping	자동분개 (Transaction) 미결관리 (Matching)	계정과목 분류예측 (개별) 내부거래 Matching	결산조정 (Closing)은 기존 회계시스템 (ERP) 내 기능 활용
Consolidation	연결회계 ; 내부거래 대사	(연결) 내부거래 Matching ; AI Connect	연결회계시스템과 연동
Reporting	외부공시 내부보고	주석자료 작성 결산보고 작성	Dart(전자공시) 시스템 데이터 내부 보고서 데이터 LLM 사전학습
Prediction	기간별 재무제표 예측 시계열 현금흐름 예측 실적/계획 평가	재무제표 예측 계정과목 금액예측 Simulation	사업계획 자료 수주/계약 정보 외부 시장정보, Dart(전자공시) 데이터
Risk Management	외부감사 내부감사	계정별/업무별 외부 회계감사 회계오류/부정적발/이상징후 감지	회계법인 AI 감사 tool 기존 Legacy 활용
D/W	데이터 및 알고리즘 관리	내부자료 수집 데이터 전처리/후처리 벡터 DB	Private Cloud IT 환경

1. AI회계의 개요

- 회계 데이터는 일반 데이터와 달리 이중기록, 기간단위, 규제이존의 특징으로 인해 정확성과 정합성 확보를 위한 정제 작업이 중요하다.

구분	회계데이터 특징	AI 적용 시 고려사항
데이터 유형	수치기반 정형데이터와 적요/품목/계정명 등 비정형 데이터 포함	<ul style="list-style-type: none"> - 머신러닝과 딥러닝 동시 적용 (비정형 데이터 처리) - 모델에 따른 데이터 전처리, 변수 선택 자동화
데이터 원천	내부 : ERP 시스템 등 외부 : BANK, HOMETAX, DART 등 (거래 원천데이터의 경우 회사 기밀사항임.)	<ul style="list-style-type: none"> - 데이터 전처리 시스템 개발 필요 - 기존 D/W 활용한 AI회계 전용 (벡터공간) DB 필요 - LLM 등 외부 API 활용 시 보안 방안
데이터 특징	시계열 데이터 생성 변수의 다양성 불균형 데이터	<ul style="list-style-type: none"> - 신규데이터(Cold start), 회계기준 변경 고려 - 코드값, KEY값, 날짜, 숫자, 키워드... 피처엔지니어링 고려 - 가중치, 데이터생성/증폭기, 알고리즘 유지보수 필요
데이터 수준	고품질의 정제된 데이터(Labeling) Logic이 내재화된 데이터 소용량의 빅데이터, 수집이 용이	<ul style="list-style-type: none"> - sLLM 개발이 상대적으로 용이함. - 요구되는 정확도가 매우 높아도 가능함. - 기존 DBMS 활용이 용이함(개발기간 단축)
데이터 생성	프로세스 기반 생성 및 집계 데이터 간 정합성 매우 중요	<ul style="list-style-type: none"> - 내부 프로세스 개선 및 정합성 체크 병행 - 기존 회계시스템(ERP) 대체는 아님. (승인 프로세스 존재)

1. AI회계의 개요

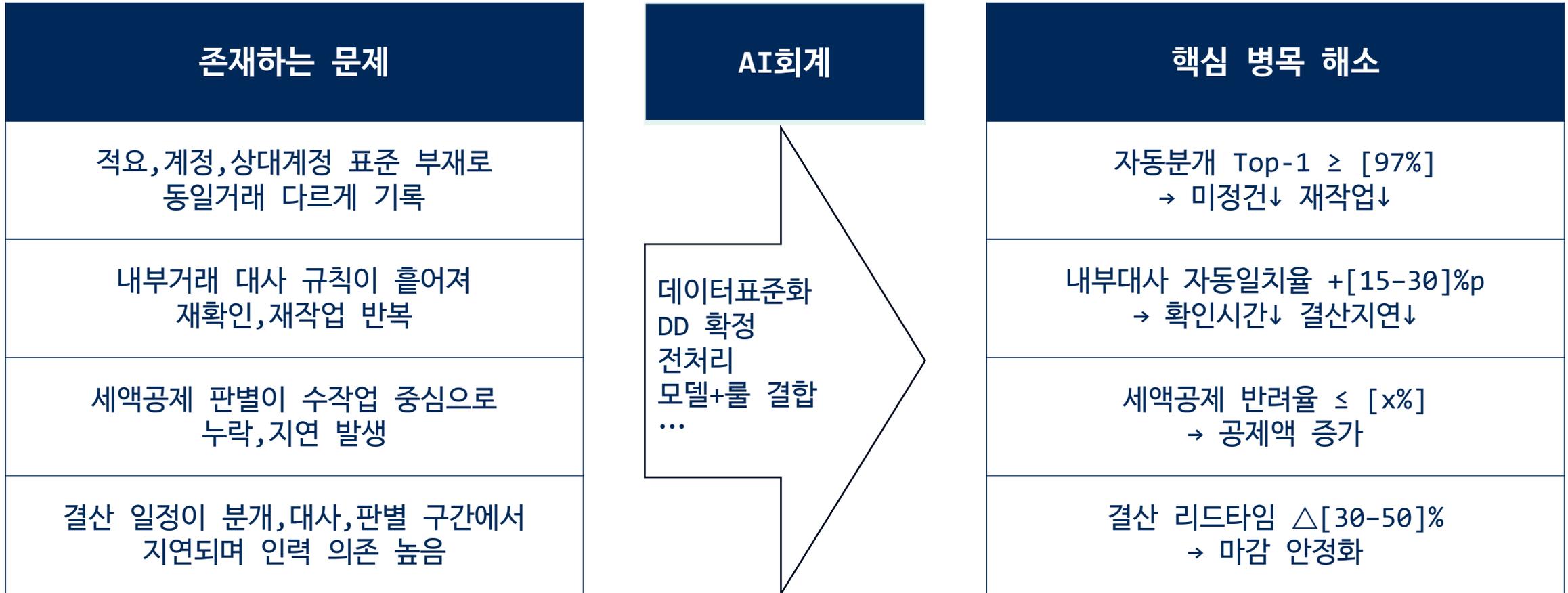


AI 회계시스템은,

거래(Transaction)단계의 전표 발생부터 거래 간 일치(Matching), 결산(Closing), 공시(Reporting)와 특정 시점의 재무제표 예측에 이르기까지 보고서 작성과 검토(감사)에 걸친 전 회계과정(Process)에서 사람(회계전문가)의 업무 개입을 최소화 한 처리과정(Process)을 구현한 시스템을 말한다.

2. 기대효과

- 수작업 분개와 수동대사는 비용, 시간적 부담이 크다. 데이터 표준화와 AI를 활용하여 동시 해결 가능하다.



3. AI 회계시스템 소개 ① 자동분개(계정과목 분류)

- Transition단계에서 예측이 가장 어려운 은행거래명세서를 자동전표 처리할 수 있는 계정과목 분류를 목표로 함

데이터 수집
<ul style="list-style-type: none"> - 4개 은행, 8개 계좌 다운로드 - 2016.1 ~ 2021.12 (5,975건)
건별 거래처, 계정코드 수기입력전표 월 결산시 오분류된 계정 수정

데이터 전처리
<ul style="list-style-type: none"> - 적요 데이터 클리닝 수행 - 결측인 변수 타변수값으로 대체
은행코드, 입출구분, 거래일자, 금액, 적요, 상대방계주, 계정, 거래처 변수사용

데이터 분리
<ul style="list-style-type: none"> - Train (2016~2020) : 4,840건 - Test (2021) : 1,135건

Result	입금 (오분류)	출금 (오분류)	전체 (오분류)
Train	2,096	2,744	4,840
Test	724 (2)	411 (15)	1,135 (17)
정확도	99.7%	96.4%	98.5%

- 오분류 분석 (1,135건 중 입금 2건, 출금 15건)

계정코	예측	결과	은행코	연	월	일	입출구	금액	원인분석	거래내용	대책	
4815	831	253	FALSE	3	2021	1	20	2	1260000	데이터부족	금융수수료	프로세스 정비 필요
4848	831	262	FALSE	3	2021	1	28	2	6000	데이터부족	금융수수료	프로세스 정비 필요
4924	253	262	FALSE	3	2021	2	15	2	176657	262가 맞음	개인카드경비	프로세스 정비 필요
4958	807	256	FALSE	3	2021	2	25	2	84652778	데이터부족	퇴직연금납입	데이터추가
5058	257	114	FALSE	3	2021	3	25	2	10000000	114가 맞음	가지급금	프로세스 정비 필요
5069	813	265	FALSE	3	2021	3	29	2	100000	데이터부족	경조사비	데이터추가
5090	253	108	FALSE	3	2021	3	31	1	55000	108이 맞음	외상매출금 회수	프로세스 정비 필요
5133	831	911	FALSE	3	2021	4	12	2	534870	데이터부족	금융수수료	데이터추가
5146	909	136	FALSE	81	2021	4	16	1	569560	데이터부족	선납세금 환급	프로세스 정비 필요
5150	253	256	FALSE	3	2021	4	19	2	495000	데이터부족	협회비	원인파악 필요?
5204	812	262	FALSE	3	2021	4	30	2	60240	262가 맞음	개인카드경비	프로세스 정비 필요
5422	812	831	FALSE	3	2021	7	8	2	4300	데이터부족	미납통행료	데이터 추가
5576	831	911	FALSE	3	2021	8	23	2	75000	데이터부족	금융수수료	데이터추가
5733	293	104	FALSE	3	2021	9	30	2	100332	데이터부족	대출금상환	데이터추가
5807	831	911	FALSE	3	2021	10	25	2	35000	데이터부족	금융수수료	데이터추가
5868	831	811	FALSE	3	2021	11	9	2	6000	데이터부족	등기수수료	프로세스 정비 필요
5869	813	811	FALSE	3	2021	11	9	2	120000	813이 아님	환풍기 필터교체	프로세스 정비 필요

희귀 유형·유사 적요에서 주로 발생.
 확정 오류 3건.

출금시 사람이 적요를 입력해
 변동성↑·유형↑로 정확도↓.
 표준 적요 도입이 핵심 처방.

3. AI 회계시스템 소개 ① 자동분개(계정과목 분류)

- 현재 홈택스(매출/매입내역)와 은행입출금내역 데이터를 업로드 하면 표준화 전처리 후 계정코드를 예측한다.
- 수집->정제->예측/매칭->검토->전표/리포트까지 단일 파이프 라인으로 자동화한다.

The screenshot displays the DAEUC&A AI Accounting System interface. On the left, a navigation menu is visible, with the '개별회계' (Individual Accounting) section highlighted in red. This section includes sub-items: '매출세금계산서' (Sales Tax Invoice), '매입세금계산서' (Purchase Tax Invoice), '법인카드' (Corporate Card), and '은행거래' (Bank Transactions). The main area shows a '자료업로드' (Data Upload) screen for '매출세금계산서' (Sales Tax Invoice). It features a search bar with '회계연월' (Fiscal Year/Month) set to '2025-07'. Below this is a table with columns for '회사코드' (Company Code), '회계연월' (Fiscal Year/Month), '순번' (Serial Number), '작성일자' (Creation Date), '송인번호' (Sender Number), '발급일자' (Issuance Date), '전송일자' (Transmission Date), '공급자 사업자번호' (Supplier Business Number), '공급자 종사업장번호' (Supplier Branch Number), '공급자 상호' (Supplier Name), '공급자 대표자명' (Supplier Representative Name), '공급자 주소' (Supplier Address), and '공급받는지' (Supplier Status). A modal window titled '매출세금계산서 자료업로드' (Sales Tax Invoice Data Upload) is open, showing '엑셀업로드' (Excel Upload) and '데이터이관플랜' (Data Migration Plan) options. It includes a '정부파일' (Official File) section with a '파일 선택' (File Selection) button and a '업로드 내역' (Upload History) section with a '데이터 이관하기' (Data Migration) button. The modal also displays a table with columns for '에러여부' (Error Status), '오류내용' (Error Content), '중복여부' (Duplicate Status), and the same columns as the main table. The interface includes a search bar, a '자료업로드' (Data Upload) button, and a '조회' (Search) button. The bottom status bar shows '100%' zoom, '0 / 0' items, and '0 to 0 of 0' items.

3. AI 회계시스템 소개 ② AI매칭(내부거래 및 미결관리)

- AI 매칭 알고리즘을 적용하여, 수동일치율을 AI매칭 일치율로 대체하여 전체 일치율 개선 및 결산일정 단축을 목표로 한다.

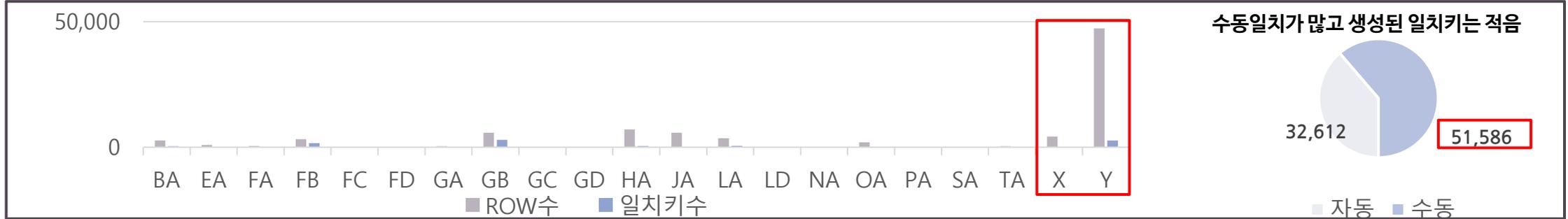
[예시]

구분	기존						개선(AI 적용)				
내부거래 데이터	최근 8년간 데이터 : 총 6,768,985 건 자동일치 29.6 %, 수동일치(금액일치) 56.5%, 수동일치(금액불일치) 6.64%, 미달/조정/제외 1.3%, 미일치 6%										
	구분	2015년	2016년	2017년	2018년	2019년	2020년	2021년	2022년	합계	평균일치율
	자동일치	210,437	202,674	225,980	257,585	281,907	313,134	263,568	244,889	2,000,174	29.55%
	수동(Y)	465,146	495,164	494,916	514,191	512,092	461,546	608,250	273,225	3,824,530	56.50%
	수동(X)	44,247	28,455	32,495	50,351	37,502	91,540	91,864	72,756	449,210	6.64%
	조정	514	32,416	577	108	7,222	2,907	7,929	37,154	88,827	1.31%
	미일치	31,902	31,902	59,107	35,450	61,490	61,444	62,502	62,447	406,244	6.00%
	계	752,246	790,611	813,075	857,685	900,213	930,571	1,034,113	690,471	6,768,985	
일치율 개선	일치율 86.1 % (자동일치 29.6 %, 수동일치(Y)- 56.5%)						일치율 88 % (프로세스 개선 후 95%) (자동일치 30%, AI일치 58%)				
결산일정 및 투입공수	D+7 ~ 14일 7일 * 5명						D+7 ~ 8일 2일 * 3명				
프로세스 개선	자동일치 - 수동일치 - 연결결산- 프로세스 개선						자동일치 - AI일치 - 연결결산 - 프로세스 개선 - AI 파라미터 조정(AI 학습기 재생성)				

3. AI 회계시스템 소개 ② AI매칭(내부거래 및 미결관리)

- 학습 데이터

2023.01 ~ 2023.09 데이터 중 일치 내역(자동, 수동) 내역을 학습자료로 활용 (84,198건(일치키 9,759개), 상대자료 생성 데이터 포함)



- AI 매칭 결과 (2023.01 ~ 2023.11) : 1월~11월 L사 누적데이터 중 Transaction 데이터

No	데이터 구분	건수	일치율	비고
1	전체데이터 중 Transaction	113,002 건		
-	조정유형 (96,97,98,99)	7,164 건		제외 (96 : 89,97 : 3,98 : 6100,99 : 972건)
2	일치율 계산 기준 건수	105,838 건		
3	컨피나스 자동일치	40,302 건	38%	자동일치
4	AI 매칭 대상데이터	65,536 건		수작업(X : 4,871, Y : 57,015) + 미 매칭(3,650건)
5	AI 매칭	57,578 건	54%	
6	총 매칭	97,880 건	92%	컨피나스 자동 + AI 매칭
7	미 매칭	7,958 건		AI 매칭 후 상대거래가 없는 데이터 393건 포함
-	매칭불가 케이스	2,313 건		상대거래가 없는 데이터(1850건), OTHER 데이터 (463건)
*	매칭불가 케이스 모수 제외 시		94.5%	

3. AI 회계시스템 소개 ② AI매칭(내부거래 및 미결관리)

- AI 매칭 결과 요약 (2023.01 ~ 2023.11) : 23년1월~11월까지의 Transaction 증감접수데이터 (96, 99 처리, 컨피나스 자동 매칭)

No	데이터 구분	건수	일치율	비고
1	전체 데이터	98,827 건		
-	조정유형(96,99)	2,166 건		제외 (96 : 126건, 99 : 2,040건)
2	일치율 계산 기준 건수	96,661 건		
3	컨피나스 자동일치	49,879 건	51%	자동일치
4	AI 매칭 대상데이터	46,782 건		컨피나스 미매칭 (46,782건)
5	AI 매칭	21,476 건	22%	
6	총 매칭	71,355 건	73%	컨피나스 자동 + AI 매칭
7	미 매칭	25,306 건		PAGE. 13 - 미매칭유형정리 참고
-	매칭불가 케이스	6,339 건		AI 매칭 전 상대거래가 없는 데이터(5,876건), OTHER데이터(463건)
*	매칭불가 케이스 모수 제외 시		79%	상대자료생성데이터 13,300건

3. AI 회계시스템 소개 ② AI매칭(내부거래 및 미결관리)

내부거래 미 매칭 유형 정리

No	미 매칭 유형
1	상대거래가 없는 데이터, OTHER 거래선 데이터 (6,339 건)
2	일방 데이터 추출 오류 (상대자료생성 : 13,300건)
3	중복자료 다수 존재 (월 단위로 자료를 받아 처리하는 과정에서 차이내역 생성 자료에 대한 삭제 처리 누락 : 2,652 건)
4	취소전표 - 발생/반제 처리 데이터 (동일한 금액이 같은 날짜에 +, - 로 발생 432건)
5	계정코드가 내부거래가 아닌 데이터 (외환차손, 선물정산손익, 파생상품 등 88건)
6	전기 미달 데이터 (전년 말 매출에 대한 1월초 매입(비용) 데이터 228 건)
7	당기 미달 데이터 (11월 중순 이후의 매출자료에 대한 매입자료 누락 1,573건)
8	거래금액 단수차이 (1 미만 차이)

AI 매칭 유형 정리

No	AI 매칭유형	유형비율
1	수익-전표번호의 합계금액과 비용-전표번호의 합계금액이 일치	59%
2	수익-거래일자의 합계금액과 비용-거래일자의 합계금액이 일치	
3	수익-적요의 합계금액과 비용-적요의 합계금액이 일치	
4	수익-거래일자의 합계금액과 비용-거래월의 합계금액이 일치 (반대 케이스도 존재)	
5	수익-거래일자(거래월)의 합계금액과 비용-건별금액이 일치 (반대 케이스도 존재)	
6	수익-적요의 합계금액과 비용-건별금액이 일치	
7	키워드를 활용한 매핑 (적요-인건비, 유지보수, LSM, PLS 등)	41%
8	수익-계정, 거래일자(n건), 법인전표번호(n개)의 합계금액과 비용 1건 금액 일치 (확률값)	
9	수익-n건 합과 비용-n건 합계금액이 일치	

3. AI 회계시스템 소개 ② AI매칭(내부거래 및 미결관리)

- 데이터 수집-> 정제-> 학습->예측/매칭-> 검토까지 단일 파이프 라인으로 자동화한다.

The screenshot displays the '데이터 생성' (Data Generation) section of the AI Accounting System. The interface includes a sidebar menu with '연결회계-내부거래' (Connected Accounting - Internal Transactions) highlighted. The main area is divided into three panels:

- 데이터 생성 실행 이력** (Data Generation Execution History): A table listing 6 data generation jobs. All are in '완료' (Completed) status.
- 서비스 진행상태** (Service Progress Status): A table showing the progress of a service job, which is '성공' (Success).
- 데이터 생성 진행상태** (Data Generation Progress Status): A table showing the progress of a data generation job, which is 'Finished'.

이력번호	처리상태	서비스명	소요시간(초)	최근 실행일시	시스템 구분
1	완료	AI 데이터관리	0.18	2024-08-27 11:22:50.048	AI_ALL
2	완료	AI 데이터관리	0.39	2024-08-27 11:14:28.990	AI_ALL
3	완료	AI 데이터관리	0.164	2024-07-29 16:52:45.660	AI_ALL
4	완료	AI 데이터관리	0.552	2024-07-29 10:45:49.496	AI_ALL
5	완료	AI 데이터관리	0.116	2024-07-23 17:10:18.288	AI_ALL
6	완료	AI 데이터관리	0.294	2024-07-23 17:08:01.471	AI_ALL

실행결과	인터페이스명	대상 서버명	jobGroupId	소요시간(초)	시작일시	종료일시	처리 메시지
성공	데이터 생성 처리 창	AI 시스템	ea8be669-8998-41de-bbat	0.108	2024-08-27 11:22:49.940	2024-08-27 11:22:50.048	데이터 인터페이스 완료

상태	Job ID	시작일시	완료일시	메시지	Job group ID
Finished	2024-08-27-11-22-51:2eb29e7c-a9d7	2024-08-27 11:22:50.020	2024-08-27 11:25:33.546		ea8be669-8998-41de-bbab-c290

3. AI 회계시스템 소개 ② 내부거래대사(AI매칭알고리즘)

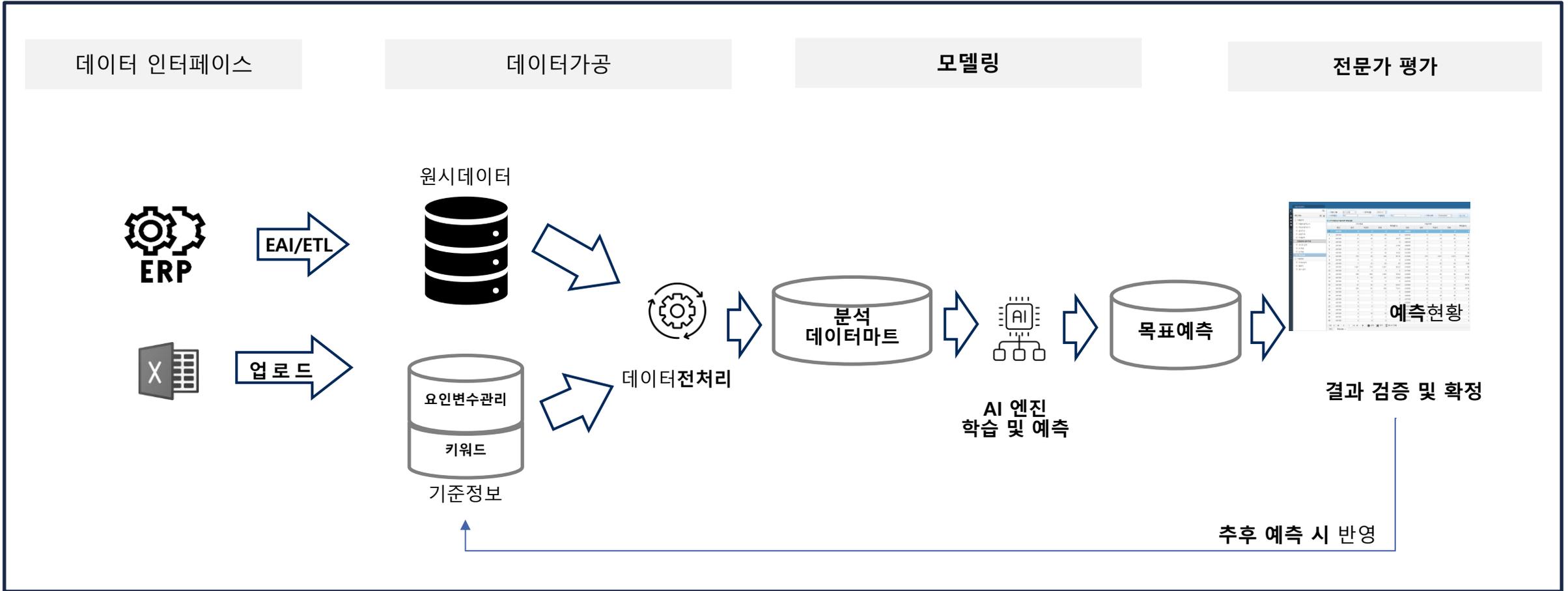
- 데이터 수집-> 정제-> 학습->예측/매칭-> 검토까지 단일 파이프 라인으로 자동화한다.

The screenshot displays the '내부거래 매칭현황 상세' (Internal Transaction Matching Details) screen. The interface includes a sidebar menu on the left with '연결회계-내부거래' highlighted. The main area shows a table of transactions with columns for '일치유형명', '일치유형', '일치번호', '수익법인', '비용법인', '거래유형', '거래일자', '거래통화', '거래금액', '장부금액', and '그룹금액'. The table lists various transactions with their respective details and amounts.

일치유형명	일치유형	일치번호	수익법인	비용법인	거래유형	거래일자	거래통화	거래금액	장부금액	그룹금액
423 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230731	USD	12,400.30	13,300,209	13,300
424 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230731	USD	4,404.96	5,638,350	5,631
425 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230528	USD	6,194.4	8,197,669	8,191
426 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230731	USD	21,139.2	27,058,176	27,051
427 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230220	USD	53,952	70,013,510	70,011
428 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	11	20230731	USD	100,411.2	128,526,336	128,521
429 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	12	20221208	USD	240,109.78	240,109.78	313,472.91
430 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,과거 연월)	AI_MLP	141	A01100	A18400	22	20230725	USD	1,414.55	1,414.55	1,846.71
431 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	190,036.6	234,011,069	234,011
432 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	13,710.09	16,882,604	16,881
433 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	111,924	137,823,214	137,821
434 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	8,741.6	10,764,406	10,761
435 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	7,720.88	9,507,492	9,501
436 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	59,793.16	73,629,298	73,621
437 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	170,788.8	210,309,328	210,301
438 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	11	20230129	USD	119,694	147,391,192	147,391
439 자기(전체, 허용오차0.05),상대(전체, 허용오차0.05),비교(년월,매입적외)	AI_MLP	183	A01100	A18400	12	20230119	USD	682,409.13	682,409.13	890,912.4
440 자기(3건Sum), 상대(2건Sum), 비교(금액)	EX_3_2	2385	A01100	A18400	11	20230930	USD	3,561.6	4,789,640	4,781
441 자기(3건Sum), 상대(2건Sum), 비교(금액)	EX_3_2	2385	A01100	A18400	11	20230930	USD	369.2	496,500	491
442 자기(3건Sum), 상대(2건Sum), 비교(금액)	EX_3_2	2385	A01100	A18400	11	20230930	USD	5,730.2	7,705,973	7,701
443 자기(3건Sum), 상대(2건Sum), 비교(금액)	EX_3_2	2385	A01100	A18400	12	20230105	USD	-5,000	-5,000	-6,521
444 자기(3건Sum), 상대(2건Sum), 비교(금액)	EX_3_2	2385	A01100	A18400	12	20221108	USD	14,661	14,661	19,140.51

3. AI 회계시스템 소개 ③세액공제 분류

- 세액공제 대상 판별 AI활용과 전문가 검증체계로 지속적인 미래전략 사업지원세액경감 적용가능



3. AI 회계시스템 소개 ③ 세액공제 분류

[테스트 데이터]

구분	대상	비대상	전체
Train 데이터	104,197	14,586	118,863
Test 데이터	27,141	3074	30,215

[정답/오답 결과]

구분	건수	건수 비율	금액	금액 비율
정답	28,405	98.4%	261,988,958,100	98.12%
오답(대상)	170	0.6%	337,205,638	0.12%
오답(비대상)	308	1.0%	4,848,768,075	1.76%

[오분류 결과]

구분	건수	건수 비율	금액	금액 비율
대상 중 오분류	1,259	4.2%	8,490,114,195	3.08%
비대상 중 오분류	73	0.2%	53,506,320	0.02%

[주요 오분류 내역]

실 공제여부	실 내용범주	예측 공제	예측 범주	wbs_text /	사유
비대상	전시물	대상	재료비	승객안전시험 계측능력 향상을 위한 장비 검교정및 센서류 구매	안전시험->전시물로 분류 케이스
대상	부품	비대상	지그	워터필터 맞세이. . (양산)	양산 포함
대상	50만원미만	비대상	지그	T.M.S MODULE (양산)	양산 포함
대상	부품	비대상	지그	스택 바이패스 밸브 맞세이 (양산)	양산 포함

데이터 특징

- 주요 feature
 - 텍스트 데이터: 텍스트, 레거시시스템변수, WBS 내역
 - 범주형 데이터: 전담부서, 소유주체, 조직 코드, 프로젝트 코드, WBS요소
 - 실수형 데이터: 금액(대상 구분 범주:50만원미만)
- 예측할 Output 설명
 - 공제여부: 세액 공제가 되는지 안되는지에 대한 0,1 분류
 - 구분: 어떤 이유로 세액 공제가 되었는지 범주형 데이터 분류
- 소유 주체가 C일 경우 공제 비대상 비율 높음 확인
 - 소유주체 그룹별 패턴 분석 필요
 - 공제내용 공통범주: 시험소모품
- 내용구분범주 상위 5개의 값이 전체 비중의 86%이상
 - 전담부서 범주가 비전담일 경우 비대상-비전담부서 확인
 - 전표유형 범주가 C3, DH일 경우 비대상- 확인
 - 텍스트 값 내 연구개발비 있을 경우 비대상- 확인
 - 문자형 컬럼 내 장갑 존재 시 비대상-시험소모품 확인
- 대상/비대상 비율 고려한 분석 데이터의 전기일자 컬럼 기준 SPLIT
 - 1월~11월 데이터 : TRAIN (대상범주 : 18, 비대상범주 : 21)
 - 12월 데이터 : TEST (대상범주 : 18, 비대상범주 : 22)
 - 차이 비대상범주: 단조, 시험장비, 출장, 취득세, 컨설팅 등

3. AI 회계시스템 소개 ③ 세액공제 분류

- 현재 세액공제 데이터 엑셀 업로드를 통한 학습 및 예측 수행 -> 집계 및 상세내역 확인 가능

The screenshot displays the '세액공제 자료업로드' (Tax Credit Data Upload) window. It includes a 'GRID 설정' (Grid Settings) panel with the following items:

- 검색
- GL계정코드
- 코스트센터
- 팀코드
- 전담부서
- 연구과제
- 연구과제명
- 소유주제
- 현지동화금액
- WBS요소
- 레벨텍스트2
- 레벨텍스트3
- WBS요소명
- 텍스트명
- VAATZ입고내역
- 세액공제여부
- 내역구분

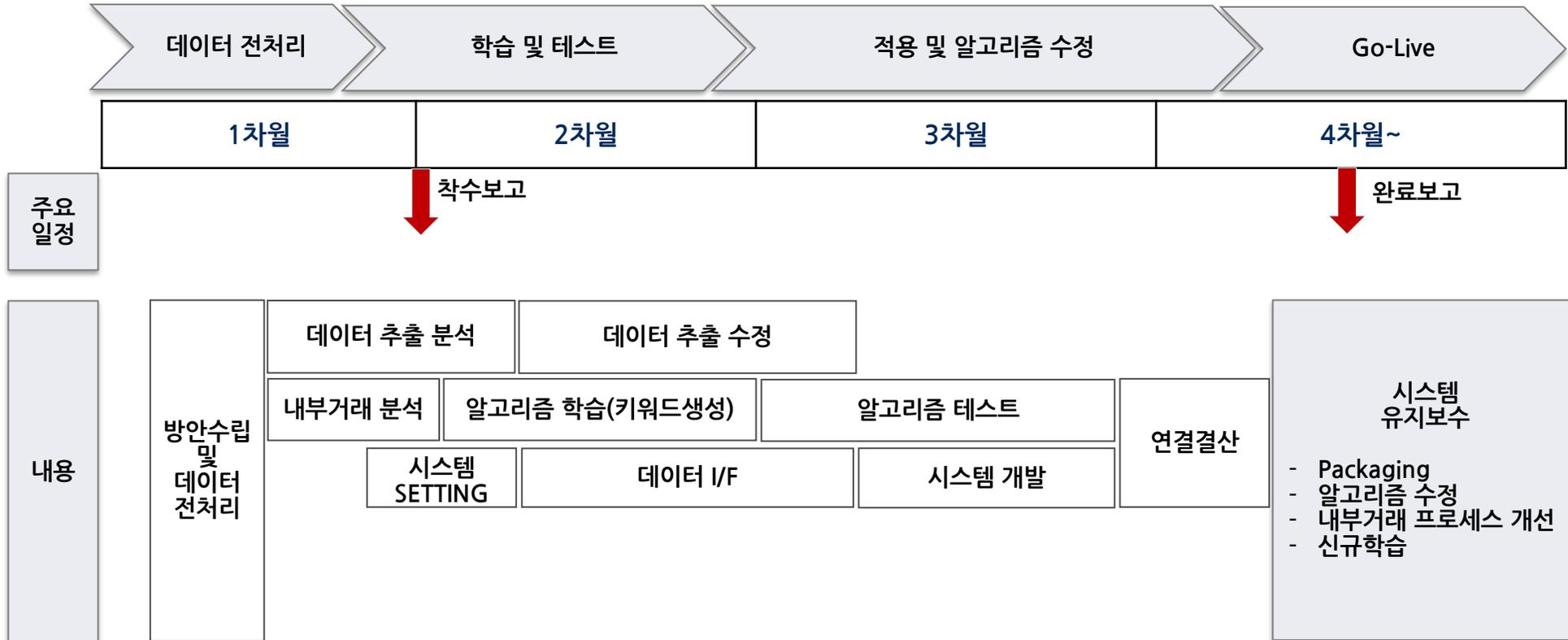
Additional settings include '헤더 행번호: 1', '표시된 컬럼' (Selected Columns), and '숨겨진 컬럼' (Hidden Columns).

번호	세액공제대상	건수
1	대상	461
2	대상	49
3	대상	1
4	대상	6474
5	대상	17420
6	대상	32
7	대상	626
8	대상	71
9	대상	384
10	대상	96
11	대상	1647
12	비대상	405
13	비대상	211
14	비대상	40
15	비대상	135
16	비대상	3
17	비대상	5
18	비대상	115
19	비대상	204
20	비대상	1
21	비대상	1
22	비대상	79
23	비대상	1583
24	비대상	35
25	비대상	110
26	비대상	27
27		118783

4. 도입 시나리오

본 프로젝트는 약 3.5개월 간 진행합니다. AI 내부거래 매칭 및 프로세스 개선을 병행합니다.

[예시]



PI컨설턴트 : 상주 1명
 AI컨설턴트 : 상주 1명, 비상주 2명
 총 6~8 M/M 소요예상

4. 도입 시나리오

구분	AI 회계서버 권장사양	
H/W	OS	Window 10 pro 이상
	VM/WSL2	Ubuntu 22.04.3 LTS (GNU/Linux 5.15.133.1-microsoft-standard-WSL2 x86_64)
	CPU	Intel Core i7~ 8Core 2.4GHz
	Memory	64GB
	Disk	SSD 1TB
S/W	Python	Python 3.11.2
	Julia	Julia 1.8.5
	Java	Jdk1.8.0_73, JDK17
	DB	PostgreSQL16
	WAS	Tomcat-9.0.43
	SSL	사내 보안정책에 따름

* 필요한 s/w는 오픈소스임

5. 향후 과제 로드맵

단기	중기	장기
데이터 수집·전처리 고도화 표준사전 v1 제작 처리시간 단축 누락·중복 항목처리 로직	연결회계 내부거래 대사 확대 법인 간 일치율 $\geq 90\%$ 결산지연 -30%	경량 sLLM 온프레미스 개발
자동분개·미결관리 파일럿 확대 Top-1 $\geq 97\%$ 자동일치율 +15~30%p	공시·보고 RAG Q&A 적용 답변 정확도 향상 응답시간 향상 외부데이터는 비식별 연계	
내부회계관리제도 강화(연결 내부거래 통제·변경관리·감사로그 불변 저장)		

6. 참고 [AI매칭 알고리즘]

- 문제 정의

- 법인 A의 데이터: (y_i^A, x_i^A) , $i = 1, \dots, N_A$

- y_i^A : i 번째 거래의 거래금액

- x_i^A : i 번째 거래에 대한 부가정보를 수치화한 것 (계정명, 적요, 거래일자 등)

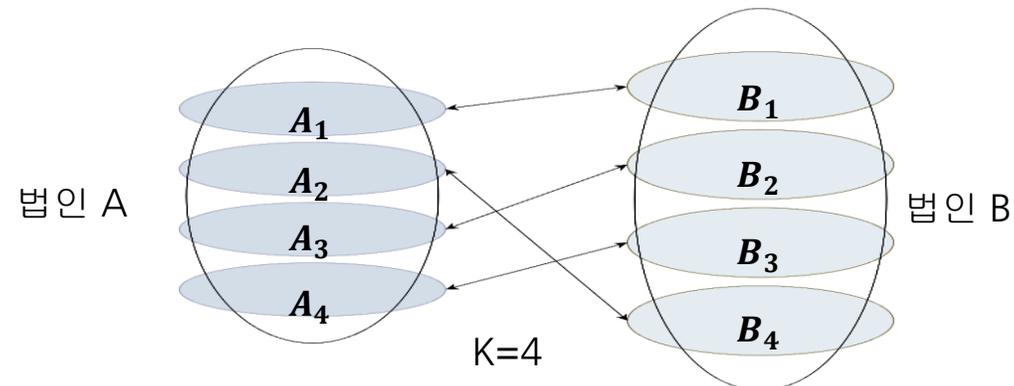
- 법인 B의 데이터: (y_j^B, x_j^B) , $j = 1, \dots, N_B$

- 원하는 결과

- 법인 A, B의 데이터가 각각 K 개의 partition, $(A_1, A_2, \dots, A_K), (B_1, B_2, \dots, B_K)$ 으로 구분되어 다음을 만족

- $\sum_{i \in A_k} y_i^A = \sum_{j \in B_k} y_j^B \rightarrow k$ 번째 partition에 속한 각 법인의 거래금액 합이 같아야 함

- 이를 만족하는 partition 중 가장 세밀하게 구분하는 partition을 찾는 것이 목적임



6. 참고 [AI매칭 알고리즘]

- 기술적 어려움

- 모든 가능한 partition에서 조건 $\sum_{i \in A_k} y_i^A = \sum_{j \in B_k} y_j^B$ 을 확인하는 것은 computational complexity가 데이터의 개수가 늘어날 수록 계산량이 폭증하므로 불가능
- 따라서 특별히 고안된 알고리즘이 필요함

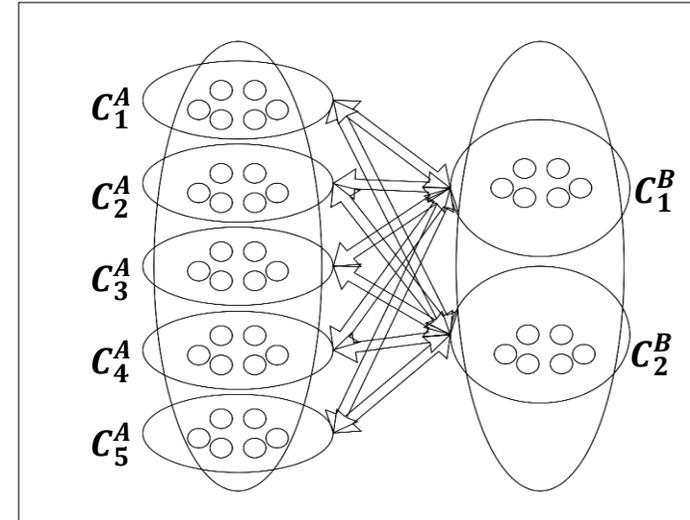
- 예제
 - 법인 A에 기록된 거래가 100건, 법인 B에 기록된 거래가 50건이라고 하자
 - 알고리즘 없이 A에 기록된 거래 중 4건과, B에 기록된 거래 중 2건의 매칭 여부 확인 위해 필요한 거래금액 합 비교 건수 (4:2 매칭이라 하자)

$$\binom{100}{4} \times \binom{50}{2} \cong \text{48억 번}$$

6. 참고 [AI매칭 알고리즘]

- 해결 방법

- 미리 법인 A와 법인 B의 데이터를 군집으로 나눔
 - 나뉜 군집들 간에 비교
 - 각 군집의 쌍 별로 거래금액 합 기반 **partition**을 찾음
- 프로세스
 - 법인 A 데이터를 k_A 개의 군집으로 군집화: $C_1^A, \dots, C_{k_A}^A$
 - 법인 B 데이터를 k_B 개의 군집으로 군집화: $C_1^B, \dots, C_{k_B}^B$
 - 모든 $i = 1, \dots, k_A$ 와 $j = 1, \dots, k_B$ 에 대해 순차적으로 C_i^A 와 C_j^B 에서 합 기반 partition을 찾음



- 예제
 - 앞의 예제와 같은 상황
 - $k_A = 5, k_B = 2$ 이고 각 군집의 크기가 각 20과 25인 경우
 - $C_1^A, \dots, C_5^A / C_1^B, C_2^B$ 로 나누고 각 군집의 쌍에서 4:2 매칭 검색
 - 즉, (C_1^A, C_1^B) 에서 4:2 매칭 검색, (C_2^A, C_1^B) 에서 4:2 매칭 검색, $\dots, (C_5^A, C_2^B)$ 에서 4:2 매칭 검색
 - 필요한 거래금액 합 비교 건수

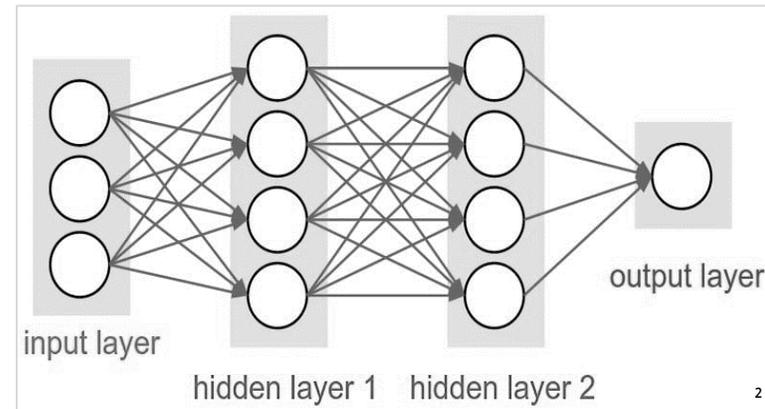
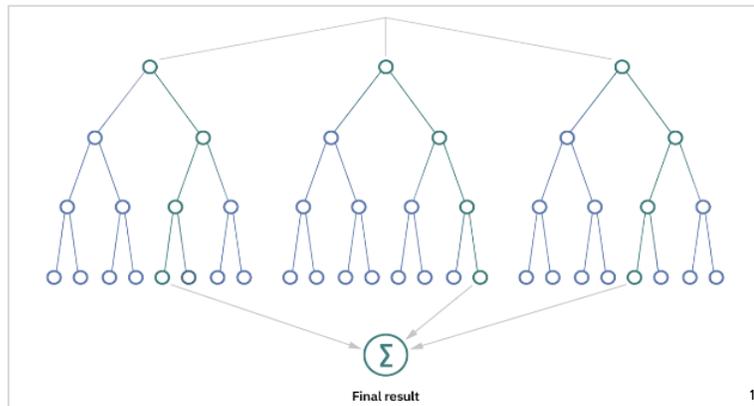
$$5 \times \binom{20}{4} \times 2 \times \binom{25}{2} \cong 1\text{천}5\text{백만 번} \rightarrow \text{약 } \frac{1}{330} \text{의 계산량}$$

6. 참고 [AI매칭 알고리즘]

- 군집화하는 방법 1.

- Matching Probability 추정

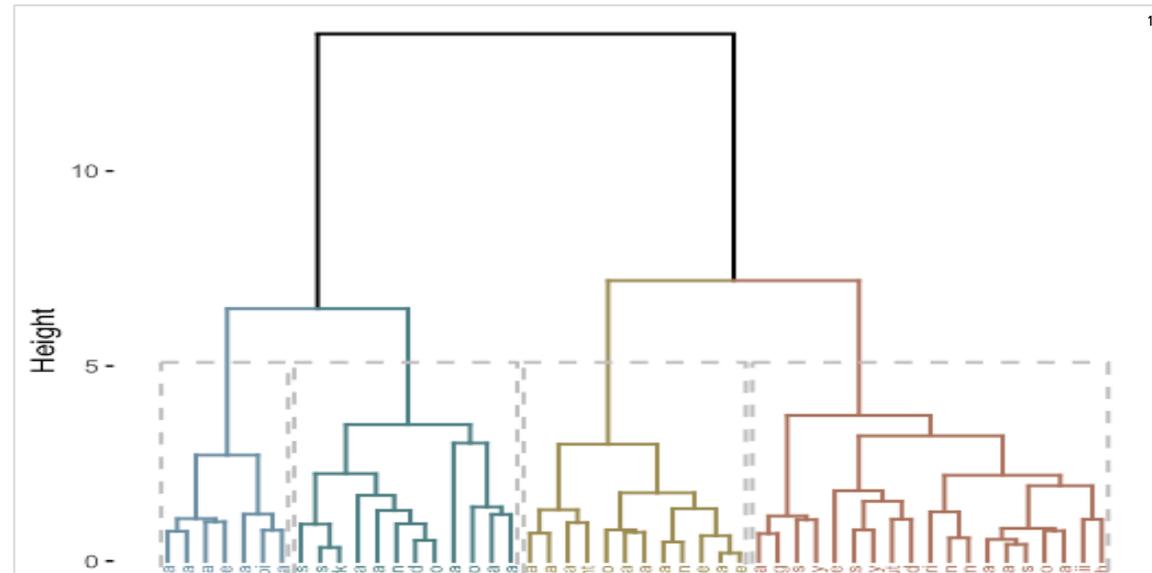
- $\rho(x_i^A)$: x_i^A 가 속하는 partition 번호
- $z_{ij}^A = I(\rho(x_i^A) = \rho(x_j^A))$: x_i^A 와 x_j^A 가 같은 partition에 속하는지 여부
- Matching Probability: $Pr(z_{ij}^A = 1 | x_i^A, x_j^A)$
 - 이를 적절한 알고리즘으로 추정
 - Ex. Random Forest, DNN 등



1. <https://www.ibm.com/topics/random-forest>
2. <https://www.bmc.com/blogs/deep-neural-network/>

6. 참고 [AI매칭 알고리즘]

- 군집화하는 방법 2.
 - 확률 기반 계층적 군집화
 - 추정된 matching probability에 계층적 군집화 알고리즘을 적용해 후보 partition을 구함
 - $Pr(z_{ij}^A = 0|x_i^A, x_j^A) = 1 - Pr(z_{ij}^A = 1|x_i^A, x_j^A)$: matching 되지 않을 확률을 x_i^A, x_j^A 두 데이터의 거리로 생각해 계층적 군집화 적용
 - 군집을 결정하는 기준은 메모리가 터지지 않는 한에서 최대 크기의 군집이 되게 함



1. <https://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering/>

6. 참고 [AI매칭 알고리즘]

- $i_A: i_B$ 매칭
 - 법인 A의 데이터 i_A 개와 법인 B의 데이터 i_B 개가 매칭되는 경우 검색
 - A 법인 거래금액 합 계산
 - 거래금액 합 계산에서 메모리가 터지지 않는 경우
 - $\{y_{i_1}^A + \dots + y_{i_{i_A}}^A \mid l_i \in \{1, \dots, N_A\}, i = 1, \dots, i_A\}$ 를 계산
 - 즉, 법인 A의 데이터 i_A 개 씩의 가능한 모든 합을 계산
 - 거래금액 합 계산에서 메모리가 터지는 경우
 - 군집화를 진행
 - $\{y_{i_1}^A + \dots + y_{i_{i_A}}^A \mid l_i \in C_k^A, i = 1, \dots, i_A\}, k = 1, \dots, k_A$ 를 계산
 - 즉, 법인 A의 각 군집 내의 데이터 i_A 개 씩의 가능한 모든 합을 계산
 - B 법인 거래금액 합 계산
 - A 법인 거래금액 합 계산과 마찬가지로 계산
 - 계산된 법인 A의 거래금액 합 목록과 법인 B의 거래금액 합 목록을 순차적으로 비교하여 매칭 진행
 - 매칭이 이루어진 데이터는 기록에서 제외시킴

6. 참고 [AI매칭 알고리즘]

- 알고리즘 프로세스
 - 각 기업에서 부가정보가 완전히 같은 데이터를 하나의 거래그룹으로 묶고 거래금액은 그 합으로 표기
 - 1:1 matching부터 시작하여 order를 높이면서 matching을 진행
 - $order = 1$
 - $i_A = 1, i_B = 1$ 인 경우 $i_A:i_B$ 매칭을 진행
 - $order = order + 1$
 - $i_A = order. i_B = 1, \dots, order$ 에 대해 $i_A:i_B$ 매칭을 진행
 - $i_B = order. i_A = order - 1, \dots, 1$ 에 대해 $i_A:i_B$ 매칭을 진행
 - 부가정보가 같아서 묶인 데이터들만을 대상으로 위의 matching알고리즘을 이용해서 세부 matching을 진행

6. 참고 [AI매칭 알고리즘]

- 알고리즘 프로세스

