



SUPREME AUDIT INSTITUTION OF INDIA
लोकहितार्थ सत्यनिष्ठा
Dedicated to Truth in Public Interest

Standard Operating Procedure for Data-led Establishment Audit



COMPTROLLER AND AUDITOR GENERAL OF INDIA

Contents

1. Purpose and Scope	2
2. Background:	2
3. Audit Approach	3
4. Data Availability:	4
5. Data Analytics	4
6. Selection of Audit entities for Field Audit	9
7. Field Verification:	9
8. Reporting:	10
9. Lesson Learnt after field verification	11
10. Responsibility Centres	11
11. Link for Useful resources	12
Annexure-I: Automation of downloading System generated Report from Web/MIS portal	13
Annexure-II: Bulk Data extraction from System generated pdfs/Reports	17
Annexure-III: Feature Engineering	22
Annexure-IV: Use of Anomaly detection tool developed by RCB&KI Jaipur	23
Annexure-V: Sample Dataset for Data-led Establishment Audit	26
Annexure-VI Reinvigorating Establishment Audit through Data Analytics in Rajasthan	27
Annexure-VII Use of AI ML in Government Sector in India	32
Annexure VIII – Lesson Learnt Report	34
Annexure IX – State wise Salary Expenditure from 2014-15 to 2022-23	36

Document Reference

SOP-EA-01/2026

Version

1.0

Issued by

Professional Practices Group, Office of the CAG of India, New Delhi

Review Frequency

Annual / As required

Applicable to

All AG (A&E) Offices, State Audit Offices, IS Wing and GA Wing of CAG Headquarters

1. Purpose and Scope

This draft Standard Operating Procedure (SOP) establishes the framework and procedures for Data led Establishment Audit to be conducted by the respective State Audit Offices. The SOP also defines responsibilities, data-sharing protocols, access facilitation mechanisms, and a structured way forward for bridging data gaps across states.

2. Background:

2.1 Trends in Establishment Expenditure of the States

Establishment expenditure of the States, comprising expenditure on salaries, pensions, and Grants-in-Aid for payment of salaries, constitutes one of the most inflexible components of State finances. The state wise salary expenditure of last 10 years is placed at Annexure-IX.

During 2023-24, committed expenditure accounted for 43.28 per cent of the revenue expenditure of the States and, when expenditure on Grants-in-Aid for salaries amounting to about ₹3.17 lakh crore (8.14 per cent of revenue expenditure) is added, nearly three-fifths of the revenue expenditure was absorbed by items of a committed nature. The object head analysis of expenditure shows that salaries, pensionary payments and Grants-in-Aid together form a dominant share of total expenditure and represent a large proportion of the combined GSDP.

The magnitude and recurring nature of these liabilities expose State finances to risks arising from even small control deficiencies in payroll and pension administration.

2.2 Common Audit Findings

CAG audit reveal systemic weaknesses including:

- Ghost employees and pensioners resulting from inadequate verification mechanisms and fragmented systems.
- Payments to deceased pensioners, continuation of salary after retirement, and duplicate payments.
- Incomplete Aadhaar seeding, absence of life-certificate verification, and lack of automated controls.
- Payroll, personnel, pension, and treasury systems operating in silos, limiting real-time cross-verification.

2.3 Data-Led Establishment Audit: Rajasthan Case Study

The Rajasthan experience demonstrates the power of data-led establishment audit. The A&E office receives fully digitised accounts and vouchers/sub vouchers from state treasuries. This enabled extraction and analysis of salary and pension data across departments and treasuries.

Key findings from Rajasthan included: approximately several instances of double payment in a Hospital and in two treasuries; and large number of instances of Dearness Allowance rates exceeding 100% in pension bills.

2.4 Leveraging E-HRMS

Most states have implemented Electronic Human Resource Management Systems (E-HRMS), either as standalone platforms or as modules within Integrated Financial Management Systems. E-HRMS generally maintains employee master data, electronic service books, and monthly payroll transactions.

3. Audit Approach

Para 4.20 of CA Guidelines 2016 state that “In case of departments/ sectors where e-governance has taken roots and transactions are being conducted in virtualised environments, digital auditing can also be adopted by the audit teams. Digital auditing facilitates looking at whole of the population for outliers or unexpected variations. Such outliers can be taken up for detailed scrutiny. Data analytical tools can be of immense help here.”

Further, at recently held AGs Conclave/Conference and senior. management meeting importance of leveraging digital platforms and IT systems for enhancing audit effectiveness was emphasized.

HQ vide letter No. 489/88-SMU/202s Dated 31 December 2025 for Annual Audit Plan 2026-27 directed that in data-led establishment audits, A&E offices would carry out data-analysis in collaboration with their counterparts in Audit and share the findings with audit offices. Subsequently, audit shall carry out substantive checks in field, if required, to clearly substantiate the audit observations. Further, based on the data analysis, PAG/AG(A&E) can bring to the attention of the Finance Secretary of the State for immediate action, instances of fraud, serious financial irregularities or major system failure. A copy of such a reference will be shared with the PAG/AG(Audit) for further follow-up.

In light of above, the Data-led audit approach has been planned around following methodologies:

- A. Rule-based anomaly detection
- B. Algorithm based anomaly detection

A. Rule-Based Anomaly Detection

Audit can analyse datasets to detect deviations from applicable financial rules. Some of the common Establishment Rules are: Basic pay for a designation will be in prescribed range; DA will be as certain percentage of Basic pay; allowances will be in defined percentage of basic pay or fixed sum according to designation /place of posting, etc.; regular salary cannot be for more than 12 months, regular salary cannot be disbursed after retirement; deductions (Income Tax, License Fee, Professional Tax, HBA) have to be as per applicable Rules,

Some other common data analytics tools and techniques which can be applied to a dataset are presence of duplicates, gap detection, trend analysis, ration analysis etc

The above exercise will help in identification of anomalous transactions such as:

- **Ghost employees and pensioners,** Payments to deceased pensioners, continuation of salary after retirement, and duplicate payments.
- **Identity Duplication:** Matching PAN, Bank Account, or Aadhaar across multiple Employee IDs.
- **Pay Scale Violations:** Salary exceeding the maximum of the assigned Pay Level.
- **Allowance Rate Validation (DA, HRA, TA)**
- **Invalid Age/Status:** Salary/Pension being drawn for employees below 18, above 60 (without extension), or post-deceased status.

By applying these rigorous filters across 100% of the transactions, the audit will move from a "test-check" of a few documents to a comprehensive verification of every single transaction against the applicable compliance requirements.

B. Algorithm based anomaly detection

While rule-based anomaly detection methodology is effective for identifying violations of certain financial rules and regulations, the Algorithm based anomaly detection using **Machine Learning (ML)** approach can be utilized to identify suspicious transactions /patterns even though they may not be any deviations from prescribed Rules. The use of AI ML for anomaly detection is an international good practice. More details can be found on website of INTOSAI¹ Working Group on Big Data and SAI20².

You can also find an article³ in Manthan magazine published by National Academy of Audit and Accounts, Shimla.

Annexure VII has details of use of AI ML in government sector in India.

By moving beyond rule-based checks to AI ML based checks, the Audit remains adaptive and catches financial irregularity that will escape traditional rule-based checks and data analytics.

4. Data Availability:

Data availability depends upon implementation of appropriate IT systems (E-HRMS, IFMS, E-Pension) as well as access to those IT systems. Most states (barring Gujarat, Arunachal Pradesh, Assam, Meghalaya, Sikkim) have implemented e-HRMS or similar system. Accordingly, the appropriate approach would be to obtain data dump periodically. In case data dump is not readily available then alternate approaches as detailed in paragraph 5.1 below can be applied.

In states where E-HRMS has **not** been implemented, VLC data or IFMS data or E-Kuber data can be utilized to find outlier (anomalous) DDOs and transactions.

The availability of data is likely to improve during 2026-27 as Government of India proposes to incentivize states for implementing Public Finance IT systems and for giving access to the CAG.

5. Data Analytics

The audit process includes the set of specific procedures, tools, and repeatable steps used to execute the audit approach. It is the "toolkit" used to gather and analyse evidence. Table 1 below gives overall status while details are given in subsequent paragraphs.

Table 1: Audit Process for Data-led Establishment Audit

Data Access	Data Fields Available	Audit Process
Comprehensive Access (VLC , IFMS, e-HRMS data) Applicable to most of the States	VLC Data /IFMS data: FY, Dept, DDO Code, District, Treasury, Head of Accounts (5-7 Tier), Voucher No/Date, Gross/Net Amt./Deductions Employee Master: Emp ID, Designation, Pay, DoB, DoA, PAN, Aadhar, DoR, PPO, Transaction Data:	Link different datasets using primary keys like transaction ID, employee ID. A. Rule based anomaly detection Apply Rule based checks to identify anomalies and seek clarifications on deviations from concerned Departments (anomalies may be clubbed Department and DDO wise to seek clarification). On same lines, A&E office can seek clarifications on same deviations from concerned treasury

¹ <https://www.iwgbd.com/files/casestudies.pdf>, www.iwgbd.com

² <https://sai20.org/storage/app/uploads/public/679/c97/4b0/679c974b0954d084882913.pdf>

³ https://cag.gov.in/uploads/subsite_whats_new/Manthan-Journal-of-Public-Sector-Auditing-and-Accounting-of-SAI-India-Volume1-Issue1-069296615a4cdb2-86088281.pdf

	Bill No/Status, Component-wise Pay (DA, HRA, TA, NPA), Deductions, Bill Paid Date.	offices as both DDO and TO are responsible for the lapse. B. AI ML based anomaly detection • Add features to above linked dataset • Apply anomaly detection to find outlier employees (using RCBKI Jaipur's Tool as detailed in Annexure-IV or through Python Library PyOD/ Orange / Knime. • PAG/AG(Audit) will conduct field verification for sample Departments and districts as detailed in para 7 below
Limited Access (VLC Data Only)	VLC Data: FY, Dept, DDO Code, District, Treasury, Sub-treasury, Head of Accounts (5-7 Tier), Voucher No/Date, Gross/Net Amt./Deductions	• Add features using RCBKI Jaipur's Tool as detailed in Annexure-III. • Identify Outlier DDOs using Outlier Analysis • Identify outlier transactions within outlier DDO • PAG/AG(Audit) will conduct field verification as detailed in para 7 below

5.1 Data Acquisition

Data acquisition involves systematic collection & extraction of data from the various available data sources. Availability of establishment related data in Audit/A&E offices could be in following manner:

- Data dump of e-HRMS and/or E-Pension database.
- Web/ MIS portal access to generate reports of employee wise transaction data.
- e-Vouchers generated from IFMS having employee wise transaction data.
- Scanned copy of salary bill with employee wise details generated through IT system

Data dump can be restored, and relevant table/key fields can be extracted with the help of CDMA wing in HQ or nearest RCB&KI or IT Audit Teams in Audit Offices. The Audit Team which might have conducted IT Audit of IFMS system in recent years can also be used for this purpose.

States which have conducted audit of IFMS recently or are in the process of conducting IFMS audit will also have data dump of IFMS data.

Automation of downloading System generated Report from Web/MIS portal can be done using Python scripts as detailed in **Annexure I**.

Bulk Data extraction from System generated pdfs/reports can also be done by customised Python scripts. Data extraction from tool developed inhouse by RCBKI Jaipur is illustrated in **Annexure-II**.

The overall goal is to secure a comprehensive dataset that includes Voucher data, DDO/Office wise employee master data, payment transactions details for further analysis.

5.2 Data Cleaning

Data cleaning is critical in AI/ML projects because poor-quality data directly leads to inaccurate models, wasted resources, and unreliable insights. In short: *garbage in, garbage out*—even the most advanced algorithms, cannot compensate for messy, inconsistent, or incomplete datasets.

Common Issues Addressed by Data Cleaning

Issue	Impact on ML Models	Cleaning Approach
Missing values	Bias in predictions	Imputation, removal
Duplicates	Skewed distributions	Deduplication
Outliers	Distorted training	Detection & handling
Inconsistent formats	Errors in processing	Standardization
Irrelevant features	Noise in learning	Feature selection

Best Practices

- **Automated tools:** Use libraries like pandas, scikit-learn, or specialized ETL platforms.
- **Domain knowledge:** Collaborate with subject experts to identify what constitutes valid versus invalid data.
- **Iterative cleaning:** Treat data cleaning as an ongoing process, not a one-time step.
- **Validation checks:** Apply statistical and logical checks to confirm data integrity.

5.3. Appending Data

To establish a robust baseline for trend analysis and outlier detection, datasets for all employees are appended to cover a continuous period of the **last three fiscal years**. By consolidating these records into a single master file, auditors can identify long-term irregularities that may not be visible in a single-year snapshot. This helps in better results (rule based or anomaly based).

5.4. Data Integration

Data integration merges disparate datasets, such as linking Drawing and Disbursing Officer (DDO) profiles with specific employee records and payment transaction data. A single, integrated table enables the auditor to see the complete "story" of a transaction.

5.5. Adding Value to Raw Data by Adding Features(Columns)

In traditional data analytics, we analyse existing dataset consisting of rows and columns for violation of applicable rules, to summarize data and to find duplicates and gaps. In machine learning, we add new columns (features) to existing dataset. These new columns /calculated fields /features add value to the existing dataset.

Addition of new columns or features depend on both domain(field) knowledge as well as technical(IT) knowledge.

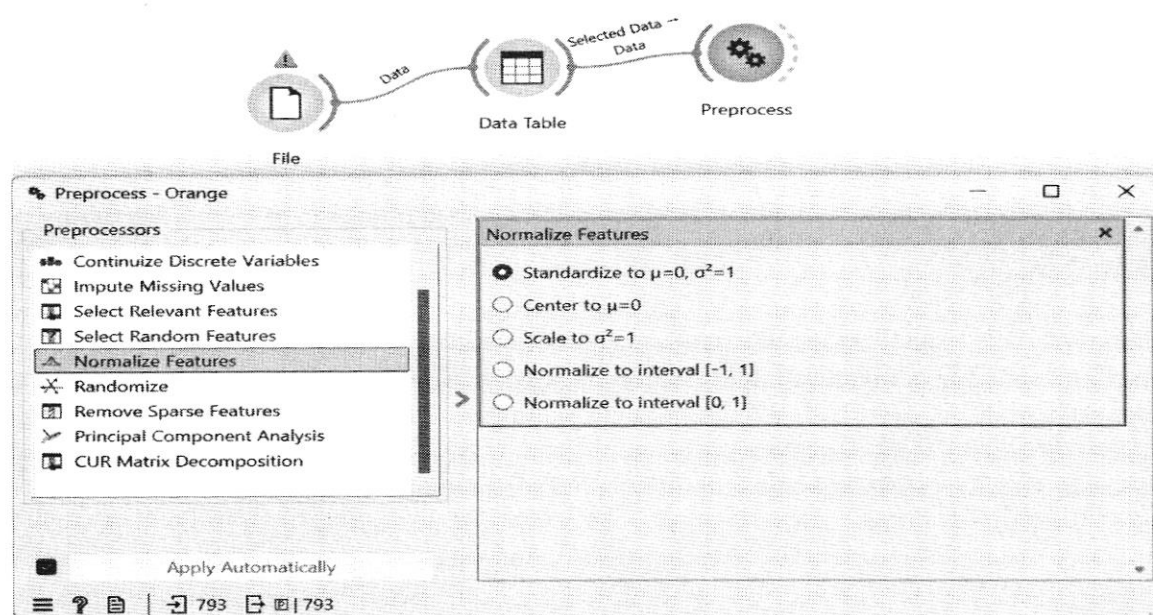
By applying mathematical logic, government rules, and temporal patterns to raw datasets (IFMS, e-HRMS, or VLC), we create new columns that specifically highlight anomalies. For example, while "Basic Pay" is just a value, the ratio of "**DA to Basic Pay**" is a feature that can immediately flag whether the system is following the prescribed government rates. This approach allows auditors to move from sampling thousands of vouchers to targeting the specific "Outliers" that represent the highest risk of fraud or leakage.

The various features which can be added are illustrated in **Annexure-III**.

5.6. Data pre-processing

Before performing anomaly detection using Orange Data Mining software /Klime /Python Library /RCBKI Jaipur tool, data preprocessing is essential to ensure accurate results. The Data Pre-processing

stage involves several activities mentioned in Figure below.



For instance, Categorical variables such as male/female need to be converted into numbers using encoding techniques, data. Features (columns in a dataset) need to be standardized (normalized in case of Isolation Forest algorithm) to bring them to a common scale. Pre-processing also involves removing of records which may have data entry errors. Data entry errors can be found by applying validation checks (age cannot be negative, or more than 110 years, or date of birth cannot be above 31,).

Additionally, feature selection or dimensionality reduction (such as removing highly correlated variables) improves model performance. Proper preprocessing enhances the reliability and accuracy of anomaly detection models.

5.7. Anomaly Detection and Outlier Analysis

Anomaly detection and outlier analysis may primarily be conducted using Python-based scripts and advanced data analytics platforms such as Orange or KNIME, in alignment with the audit approach outlined in the previous section. The specialized tool developed by RCB&KI Jaipur may also be utilized, wherever considered appropriate. The process of detecting anomalous & outliers transaction using Inhouse developed tool is detailed in **Annexure-IV**. By identifying outliers /anomalies, the audit can focus resources on the most suspicious transactions that deviate from the normal behaviour of similar groups.

5.8. Rule based anomalies

Beyond algorithmic detection, auditors must apply specific business rules based on Government Orders (G.O.s) to identify definitive compliance breaches. Drawing from the Rajasthan experience, the following rules should be automated for 100% population screening:

Duplicate Disbursement Checks: Identify "Identity Duplication" by matching PAN, Bank Accounts, or Aadhaar across multiple Employee IDs. For example, verify if the same person received pay via two different billing systems (e.g., Pay-manager vs. IFMS 3.0).

Ghost Employee & Pensioner Detection:

- **Post-Retirement/Superannuation:** Flag any salary drawn after the official date of retirement.

- **Invalid Age:** Flag transactions for employees recorded as below 18 or above 60 without an official extension.
- **Deceased Status:** Cross-link pension data with life-certificate databases to stop payments to deceased pensioners.

Allowance & Pay Scale Validation:

- **Ceiling Breaches:** Flag when Basic Pay breaches the range assigned in Pay Scale/Level.
- **Rate Discrepancies:** Validate DA, HRA, and Mess allowances against current G.O.s. (e.g., HRA exceeding the 8-27% slab or DA rates incorrectly applied from older Pay Commissions) .

Commutation Regulatory Breaches: Ensure commutation payments do not exceed the statutory ceiling and are not incorrectly deducted from Family Pension heads.

6. Selection of Audit entities for Field Audit

Anomalies found through AI ML needs to be verified in the field. These anomalies can be grouped on the basis of Department, district, treasury and DDO as integrated transactional(payment) data is supposed to have all above fields.

The selection of audit units is transitioned from traditional random sampling to a **Risk-Based Stratified Approach**. Under this methodology, districts and Drawing and Disbursing Officers (DDOs) are selected by PAG/AG(Audit) based on the density and severity of anomalies detected during data interrogation. This ensures that audit resources are surgically deployed to areas with the highest probability of financial leakage or systemic failure.

Stage I: District and Treasury Selection

In the first phase, audit units are prioritized at the macroscopic level by aggregating individual anomalies to identify high-risk geographical hubs. The selection is executed through the following steps:

1. **Aggregation of Anomalies:** The total number of anomalous transactions is calculated and aggregated for each District or Treasury based on the flagged dataset.
2. **Ranking of Districts:** All Districts and Treasuries are ranked in descending order based on the absolute count of detected anomalies.
3. **20%** of the highest-ranked districts (subject to minimum of two and maximum of 10) in the audit universe may be selected for field audit. While selecting district/treasury at least 50% should be from remote or border areas.

Stage II: Selection of DDO

Within sample district, attempt may be made to verify 100% anomalies. However, if examination of 100% anomalies involves more than 33% of DDOs in the district then concerned PAG/AG(Audit) may apply his/her discretion in selection of the DDOs depending upon available mandays.

7. Issues for examination during Field Verification:

During field verification, the Audit Team is expected to conduct comprehensive and 360-degree assessment in respect of outlier (anomalous) employee or transaction. Below is an implementation checklist for conducting an effective Field Verification.

1. Core Personnel & Pay Fixation

- **E-service Book:** Check master data (name, father name, Date of Birth, Date of Joining, family details, Aadhar Number, Family ID, if applicable) for correctness.
- **Service Books:** Verify that service books are maintained accurately, updated, and kept in the safe custody of the Head of Office. Check for annual entries, verification of service, and attestations.
- **Pay Fixation:** Verify the initial pay fixation upon appointment, promotion, or re-fixation under the Pay Commission rules. Ensure increments are granted on the correct due dates and correctly calculated.
- **Appointments/Promotions:** Check for valid sanction orders. Ensure the appointments adhere to the prescribed recruitment rules (RRs) and that all necessary vetting (e.g., vigilance clearance, medical fitness) was completed.
- **Leave Records:** Verify leave accounts. Ensure that leave is sanctioned by the competent authority, correctly debited, and that leave encashment is paid as per current entitlements.

2. Allowances & Entitlements

- **Dearness Allowance (DA):** Ensure that DA rates applied align with the latest orders issued by the Ministry of Finance.
- **House Rent Allowance (HRA):** Verify the classification of the city/town against HRA percentage entitlement. Ensure certificates (for non-availability of government accommodation) are obtained where required.
- **Transport Allowance:** Check entitlement based on pay levels and ensure it is not drawn during periods of unauthorized absence, leave, or tour where a separate daily allowance is claimed.
- **Medical Reimbursement:** Ensure claims are supported by proper prescriptions and are within the scope of the Central Services (Medical Attendance) Rules.

3. Deductions & Statutory Payments

- **GPF/NPS:** Verify that subscriptions are deducted correctly as per the employee's choice/nomination and credited to the appropriate funds.
- **Income Tax (TDS):** Check that tax is deducted at source based on the declared investment and salary projections; verify that these are deposited into the government account within the stipulated time.
- **Professional Tax:** Ensure compliance with the respective State Government's professional tax slabs.

4. Retirement & Termination

- **Pension Papers:** Review the promptness and accuracy of processing pension papers. Ensure that all dues (Government and non-Government) are calculated and adjusted before final settlement.
- **Full & Final Settlement:** For resignations or deaths, ensure no-dues certificates are obtained and that retirement benefits (Gratuity, Leave Encashment, GIS) are disbursed in accordance with the rules.

5. MACP

- **Service & Eligibility Validation:** Verify the "Regular Service" length in the Service Book to ensure non-qualifying periods (e.g., ad-hoc service, *dies-non*, or EOL without medical certificate) were excluded from the 10/20/30-year calculation.
- **Performance & Discipline Scrutiny:** Cross-check the Departmental Screening Committee (DSC) minutes against APAR ratings and the Penalty Register to ensure the benefit wasn't granted during a currency of penalty or despite "Below Average" performance.
- **Fixation & Hierarchy Audit:** Recalculate the 3% notional increment and verify that the pay fixation was restricted to the immediate next higher Pay Level in the hierarchy, ensuring no "double benefit" was given for the same grade.

8. Reporting:

The reporting framework for Data-Led Establishment Audits consists of:

- a. Issue of Audit Note /Observation from HQ to all Departments in respect of rule-based anomalies by Audit Office. The Audit Note /observation may be prepared theme wise (incorrect basic pay, DA, allowances, deductions or double payment or salary paid after retirement/death). Simultaneously, A&E office may seek clarification on said anomalies from treasuries.
- b. Field audit teams may issue audit memos and seek clarification from DDOs in respect of anomalies that they have verified during field visit.

- c. Based on responses to audit memos issues during field visit, Department wise Inspection Report may be prepared and issued to the Head of the Department for response /follow up action. The said Inspection Report should find root causes for anomalies and give recommendations to prevent recurrence of similar lapses in future.

9. Lesson Learnt after field verification

Field verification provides a measurable assessment of the success rate of anomaly detection techniques used in the data-led audit. By comparing flagged outliers with actually confirmed irregularities, the Audit Office can evaluate key indicators such as confirmation ratio (true positives), false positives, and missed cases. A high confirmation rate indicates strong model precision, while deviations highlight the need for recalibration of thresholds, parameters (e.g., contamination rate in Isolation Forest), and rule-based checks. The feedback from field verification should be systematically incorporated into model retraining and feature refinement to improve accuracy, sensitivity, and overall effectiveness of anomaly detection in subsequent audit cycles.

Audit Teams engaged in Establishment Audit are expected to submit a Lessons Learnt in prescribed format to the Data Analytics Cell of concerned Audit office and to RCB&KI for institutional knowledge, growth and improvement.

10. Responsibility Centres

The following responsibility matrix defines the roles of key stakeholders in implementing this SOP:

Stakeholder	Key Responsibilities
Audit Offices	1. Primary execution of the audit, field verification of flagged anomalies. 2. Specify data and voucher requirements at the start of each audit cycle. 3. Anomaly detection using Rule Based Checks & ML Algorithm. 4. Field verification of anomalous/outlier/suspicious transactions/DDOs. 5. Issue of department wise Inspection Report to Head of Department. 6. Issue of Inspection Report to concerned unit/DDO 7. In addition to salary, Audit Office will also check pensionary payments in case such payments are taking place through banks
Data Analytics Cell of Audit Office	1. Conduct rule based and AI ML based anomaly detection 2. Review of Lesson learnt report and suggest improvement in features and anomaly detection
AG (A&E) Offices	1. Provide VLC data/IFMS /e-HRMS data to Audit offices. 2. Seek clarification on rule-based anomalies from the treasury 3. Conduct rule based and AI ML based anomaly detection for pensionary payments where treasuries and not banks are making payments 4. Bring to the attention of the Finance Secretary for immediate action, instances of fraud, serious financial irregularity and major system failure/weakness, noticed from the data analysis.

Stakeholder	Key Responsibilities
AG (A&E) Offices With Pension / GE Wing	- Capacity Building of Audit offices regarding establishment matters, pay fixation, - Sharing of relevant GOs/orders/Guard files.
IS Wing (O/o CAG)	- Ensure availability of requisite IT infrastructure (scanner, Server, GPU) to A&E and Audit offices. Common for all Data-led audits (social sector, works audit and establishment audit) - Provide technical knowhow and handholding for retrieving data from data dump
KCB Wing & All RCB&KI	- Capacity Building in AI/ML, Python & OCR technology, Database Import/restore, Data integration, Data analytics & web Scrapping. - Capacity building in use of Anomaly detection & outlier Analysis - Development of STM & Case studies for Data led Establishment Audit.
State Finance Departments / NIC	- Grant read-only accesses to e-HRMS Portal to Audit and A&E offices. - Provide data dumps of e-HRMS database at agreed frequencies;

11. Link for Useful resources

Data Extraction In-house Tool of IA&AD	https://workdrive.mgovcloud.in/file/0gc1hb19c53a735cd4a08a80362bca43e55c2
Anomaly Detection In-house Tool of IA&AD	https://workdrive.mgovcloud.in/file/0gc1h9744e683e57f4fe48da80398250b885
INTOSAI Working Group on Big Data	https://www.iwgbd.com/files/casestudies.pdf , www.iwgbd.com
India AI Summit	https://impact.indiaai.gov.in/

Annexure-I: Automation of downloading System generated Report from Web/MIS portal

What is Web Scraping in Auditing?

Web scraping is the process of automatically collecting information from a website using a computer program. In auditing, web scraping can be used to collect publicly available information from the auditee entity's website for verification, cross-checking, and analytical review. It helps auditors automatically gather relevant data instead of manually copying information page by page.

Why Auditors Use Web Scraping?

- Collect product or service details published online
- Extract pricing information
- Review announcements and press releases
- Verify regulatory disclosures
- Compare website disclosures with financial statements
- Monitor changes in website content over time

How Web Scraping Works in Audit Context

1. Access the auditee's website.
2. Download webpage content.
3. Identify relevant audit information.
4. Extract required data in structured form.
5. Store data as audit evidence (Excel/working papers).

Role of Python in Web Scraping

Python is widely used because it is simple, readable, and has powerful open-source libraries. It can efficiently collect, process, and export data into Excel for audit documentation.

Common Open-Source Python Libraries

Open-Source Library	Purpose
Requests	Connects to websites and downloads webpage content
BeautifulSoup	Extracts text, tables, and disclosures from webpage structure
Scrapy	Advanced framework for large-scale scraping
Selenium	Automates browser actions for dynamic websites

Advantages for Audit Work

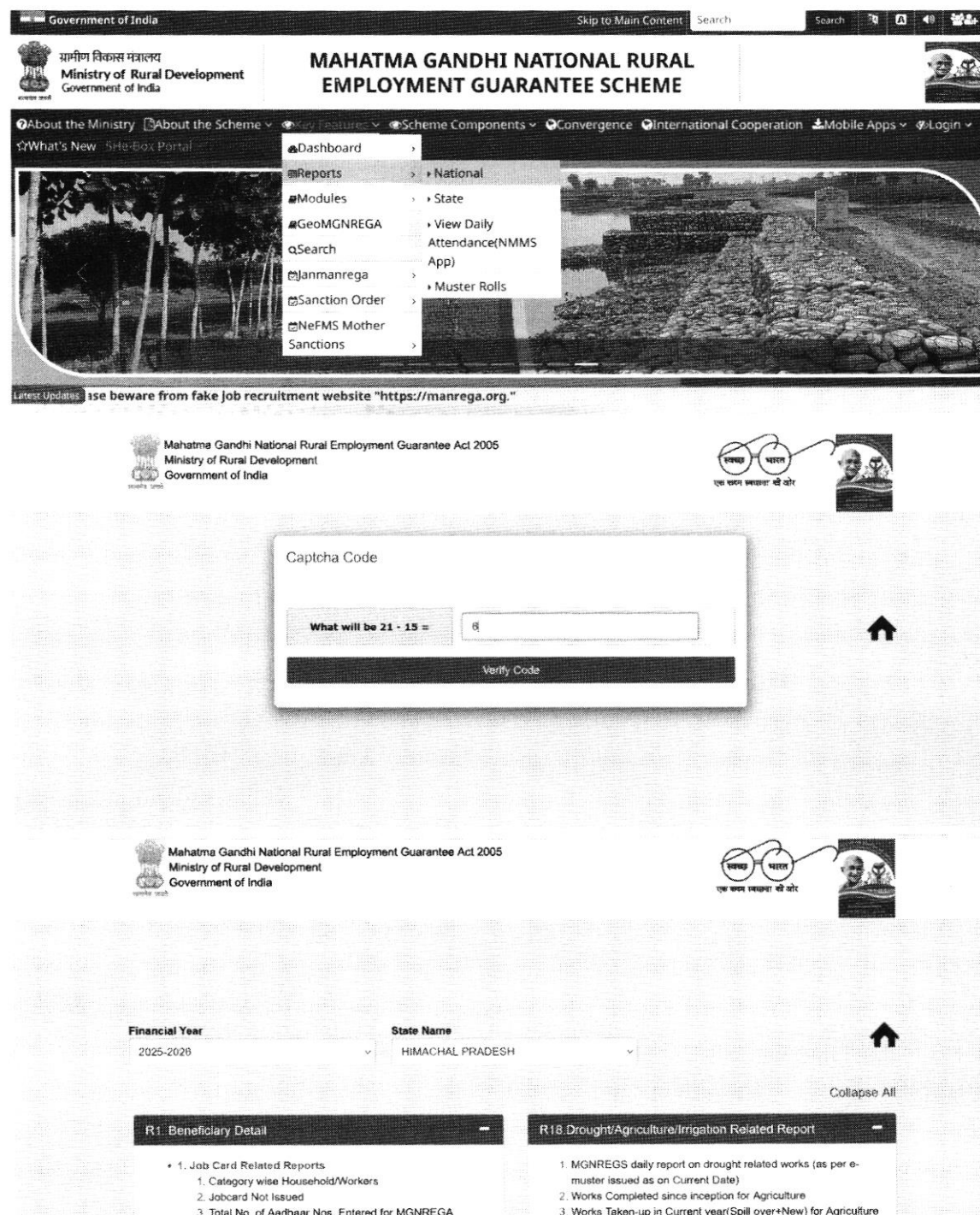
- Saves time and manual effort
- Improves efficiency of audit procedures
- Supports analytical review
- Provides documented audit evidence
- Reduces human error

Legal and Ethical Considerations

Auditors should collect only publicly available information, follow website terms and conditions, avoid restricted areas without authorization, and ensure data is used strictly for audit purposes.

Case Study on Automation Process of Website

This is a case study in which an automation process was implemented for the NaregaSoft website of the state of Himachal Pradesh using Python libraries.



Government of India

Skip to Main Content

Search

श्रीमती विकास मंत्रालय
Ministry of Rural Development
Government of India

**MAHATMA GANDHI NATIONAL RURAL
EMPLOYMENT GUARANTEE SCHEME**

What's New

Home-Box Portal

Dashboard

Reports

Modules

GeoMGNREGA

Search

Janmanrega

Sanction Order

NeFMS Mother Sanctions

National

State

View Daily Attendance(NMMS App)

Muster Rolls

Latest Updates

Use beware from fake job recruitment website "https://manrega.org."

Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India

कर्मकाय भारत
एक सपना हमकाय से जोड़

Captcha Code

What will be 21 - 15 =

Verify Code

Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India

कर्मकाय भारत
एक सपना हमकाय से जोड़

Financial Year

2025-2026

State Name

HIMACHAL PRADESH

Collapse All

R1. Beneficiary Detail

1. Job Card Related Reports

1. Category wise Household/Workers

2. Jobcard Not Issued

3. Total No. of Aadhaar Nos. Entered for MGNREGA

R18. Drought/Agriculture/Irrigation Related Report

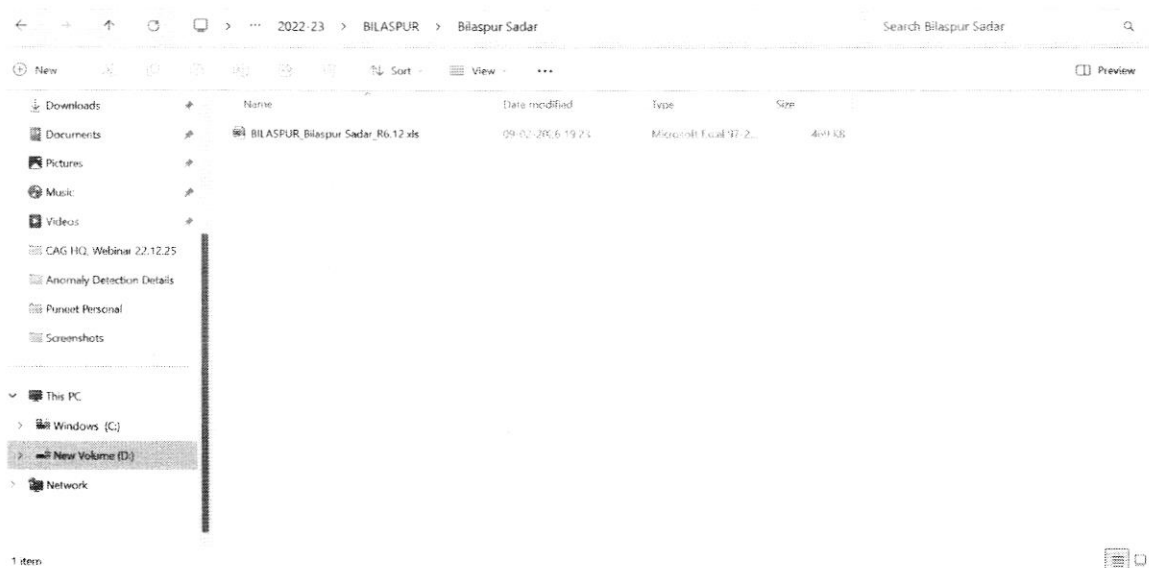
1. MGNREGS daily report on drought related works (as per muster issued as on Current Date)

2. Works Completed since inception for Agriculture

3. Works Taken-up in Current year(Spill over+New) for Agriculture

Sl. No.	Task	Completion Status
1	Work Status	
2	Yearly Work Completion Rate	
3	Master Work Category wise Analysis	
4	Spill Over Works	
5	Details of Completed work by implementing agency	
6	No. Of Works Under Convergence Entered in MIS	
7	Report on individual Household Toilets	
8	Work Category wise No. Of Works Under Convergence Entered in MIS	
9	MGNREGS daily status as per e-muster issued	
10	Expected Outcome VS Actual Outcomes	
11	Household From vulnerable communities Stopped coming for MGNREGA Works	
12	Dynamic Report for Monitoring and details of works	
13	Dashboard for incomplete works	
14	Physical completion pending for final completion	
15	Map of IAY Houses Status	
16	Railway Related Works	
17	Status of works shared to SECURE/BHUVAN	
18	Worksite Photograph & Drawing Report	
19	Construction of Biogas plant for individual	
20	Comprehensive Report on Plantation Maintenance, Progress, and Green Infrastructure Management	
21	Report on individual Benefit Works (IBS)	
22	Tax Reconciliation	
23	Monthly Status Report on Completed Works	
1	Flagged Work Status For GEO Tagging	
2	Monitoring Report for Asset Id	
3	Monitoring Report for GeoMGNREGA Training	
4	GeoMGNREGA Phase I Comprehensive Reports	
5	State-wise Janamranga user report	
6	GeoMGNREGA Phase II Comprehensive Reports	
1	eTicket Status Report	
2	State wise eTicket Status Report	
3	Search eTicket	
1	Work Completion status	
2	Master Circular/New Asset Register Status/Seven Register Status	
3	Progress of DUGWELLS and WELL	
4	Verification of Job cards	
5	SECC Mapping Status for Landless Casual Labour	
6	Individual Asset Creation	
7	Work Structure wise Physical & Financial Report	
8	Status of work by category	

No.	District	Block	Panchayat	Work Start Fin Year	Status	Work Code	Master Work Category Name	Work Category per new work creation Module	WORK TYPE as Type	Sanction Amount (in Rs.)	Sanction Amount (in Rs.)	Amount in cur year		
1	Bilaspur	BALH	Sadar BHALWANA	2026 Ongoing	IFAY/55/268	1301001001/ CONSTRUCTION OF PLAY-G HOUSE FOR INDIVIDUALS - ALONG WITH PROTECTION WALL NEAR HOUSE OF MAST RAM S/O UDHO RAM VILL TAHER CP BALH BALTHAWANA	B-INDIVIDUAL CONSTRUCTION OF PLAY-G HOUSE ASSETS FOR VULNERABLE SECTIONS ONLY FOR INDIVIDUALS IN PARAGRAHS S)	Works on construction of individual play-g house building for individuals (Category) Individuals	Levelling/shaping of land fallow / Wasteland / Individuals	Levelling/shaping of Gram	30400	0	30400	27200
2	Bilaspur	BALH	Sadar BHALWANA	2025-2026 Ongoing	IF/8000285326 1301001001/ LAND DEV HP137991208	NO LONG WITH ALONG WITH PROTECTION WALL NEAR HOUSE OF MAST RAM S/O UDHO RAM VILL TAHER CP BALH BALTHAWANA	B-INDIVIDUAL LAND DEV HP137991208	Works on Levelling/ shaping of land fallow / Wasteland / Individuals	Levelling/shaping of Gram	30616.89	18422.23	49039	29760	
3	Bilaspur	BALH	Sadar BHALWANA	2025-2026 Ongoing	IF/8000285337 1301001001/ LAND DEV	ASSETS FOR INDIVIDUALS	B-INDIVIDUAL ASSETS FOR INDIVIDUALS	Works on Levelling/ shaping of land fallow / Wasteland / Individuals	Levelling/shaping of Gram	51512.43	48429.15	99942	51340	



Conclusion

Web scraping is a valuable tool for auditors to efficiently gather publicly available information from an auditee's website. When used responsibly, it strengthens audit verification, documentation, and analytical procedures.

PDF Table Extractor

User Guide

For System-Generated PDFs Only
Puneet Sharma, AAO/RCB&KI, Jaipur

IMPORTANT — System-Generated PDFs Only

This tool works exclusively with system-generated (digitally created) PDFs — such as government reports, vouchers data mainly for tabular type data.

It will NOT produce reliable results on scanned documents, photographed pages, or image-based PDFs. For scanned files, LLM must be used first.

1. Compatible Document Types

This tool extracts tabular data from PDFs that were **generated directly by software**. Scanned/image PDFs are not supported.

1.1 Supported Document Categories

Document Type	Examples	Key Characteristics
Auditee Data	Auditee Entity Data stored in system Generated PDFs	Wide landscape tables, split across pages
Salary & Pension Vouchers	Monthly pension bills, salary registers, Pay-in-slips	Serial numbers, rows spanning multiple pages
Financial & Budget Documents	Treasury statements, expenditure registers, budget PDFs	Numeric-heavy, multi-table pages
Database / ERP Exports	PFMS, HRMS, Vouchers Data, other Data saved in pdfs	Machine-perfect alignment, consistent columns

1.2 Supported Table Structures

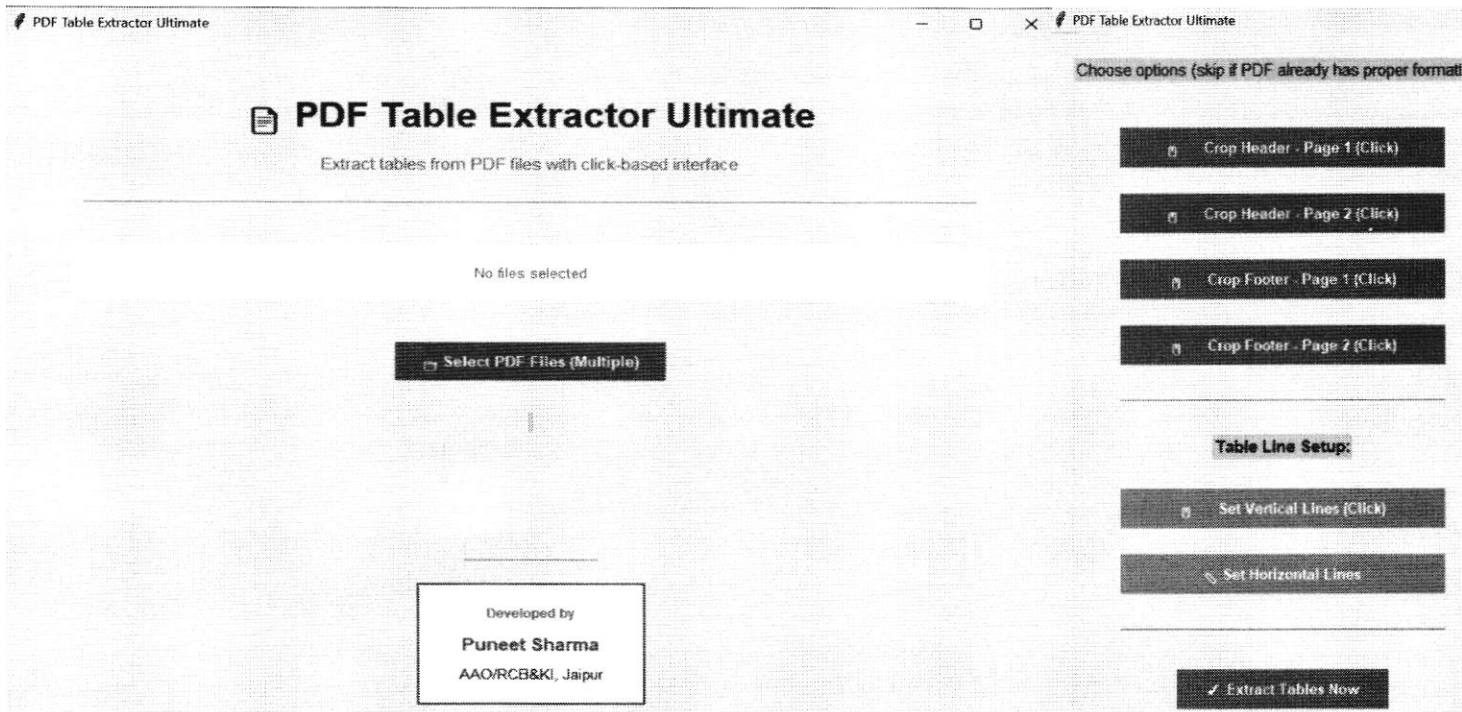
- **Simple single-page tables:** Rows and columns separated by drawn lines, one table per page.
- **Wide tables split left-right across two pages:** Same 33 rows appear on page 1 (first half of columns) and page 2 (second half). Common in government Auditee Entities.
- **Multi-page continuous tables:** Serial numbers run 1, 2, 3 ... continuously across pages. A row can start at the bottom of one page and finish at the top of the next (e.g., pension bills).
- **Multiple tables per page:** Each table extracted separately with its own table number in the output.

TIP — How to Verify Your PDF is System-Generated

Open the PDF. Try to click and drag to select text. If text highlights in blue and copied, the PDF is system-generated and compatible.

If clicking selects only an image area with no highlighted text, the file is scanned and requires OCR before use.

This is the interface of the **In-house built Data Extraction Tool** for the IA&AD Department.



2. Step-by-Step Procedure

Follow all steps in order. Each step builds on the previous one. Do not skip steps.

1

Launch the Tool

Run the file `pdf_table_extractor_v12.py` using Python 3.

A window titled "PDF Table Extractor" will appear with five action buttons arranged top to bottom.

2

Select PDF Files — Click "Select PDF Files"

A file browser opens. Navigate to your PDF folder.

Hold Ctrl to select multiple files at once. All selected files will be processed in one run.

The tool generates one CSV output file per PDF, saved in the same folder.

3

Crop Page Header — Click "Crop Header Page 1" and "Crop Header Page 2"

A page preview opens with a horizontal cursor line that follows your mouse.

Click just above the first data row to mark the crop point — this removes title text and column-heading rows from extraction.

If both page 1 and page 2 headers have the same height, set the same crop for both.

Skip: If no header crop is needed, click Back without clicking on the page. The tool still handles the column-number row (1 2 3 4 ...) automatically.

4

Set Vertical Lines — Click "Set Vertical Lines" (Optional)

The tool auto-detects column boundaries from the PDF's own drawn vertical lines in most cases. Skip this step first.

Only use this step if columns appear merged in the output. A page preview opens — click along the top to place column dividers manually.

- First Enter Number of lines which you want to draw on the page (Tip- No. of column +1)
- Draw all vertical lines one by one by clicking on suitable place.

5

Set Horizontal Lines — Click "Set Horizontal Lines"

This opens a sub-menu. Select the option that matches your document type:

Option	When to Use
Detect Serial Number	Pension bills, registers, and any table with a serial number column (1, 2, 3 ...). Handles per-page column widths and wide two-page tables automatically.
Line by Line	Simple tables where each printed line is one data row and spacing is consistent.
Set Manually	Tables with no drawn lines and irregular spacing. Click on the preview to place row dividers by hand.

6

Extract Tables — Click "Extract Tables Now"

The tool processes all selected files. A progress bar and live page counter (e.g., Page 4/8) update in real time.

When complete, a message shows the output file path. Each PDF produces one CSV.

The CSV columns are: PDF_Name, Page, Table, Column_1, Column_2 ... Column_N

7

Save Configuration — Click "Save Settings" (Optional but Recommended)

After a successful run, save your settings to a JSON file on your Desktop.

Use "Load Settings" next time to restore all crop and table settings instantly — ideal for processing the same document type repeatedly.

NOTE — Output File Location

The CSV is saved automatically in the same folder as the input PDF, with `_extracted.csv` appended to the file name.

Example: `District_Report.pdf` → `District_Report_extracted.csv`

Columns: `PDF_Name` | `Page` | `Table` | `Column_1` | `Column_2` | ... | `Column_N`

3. Data Extraction Summary

3.1 Quick Reference — Buttons at a Glance

Button	Purpose
Select PDF Files	Open file picker to choose one or more PDFs. Multiple files processed in one run.
Crop Header Page 1/2	Click on the page preview to mark where data begins. Removes title and column heading rows.
Set Vertical Lines	Manually place column dividers. Skip if auto-detection works (it usually does).
Detect Serial Number	Automatic mode for tables with a serial number column. Handles per-page column layouts automatically.
Line by Line	Extracts each printed text line as one row. Best for simple evenly-spaced tables.
Extract Tables Now	Starts extraction. Live progress bar shows page count.
Save Settings	Saves current crop and table settings to a JSON file for reuse.
Load Settings	Restores previously saved settings instantly.

SUMMARY — The Complete Two-Stage Workflow

Stage 1 — Extract: Use this tool to convert PDF tables into a raw CSV. Column detection, serial tracking, header removal, and multi-page handling are all automatic.

Stage 2 — Clean: Open the CSV in Python (pandas) to remove separators, strip commas, filter footer notes, convert data types, and rename columns. Save the result as your final analysis-ready dataset.

Input File – Pension Data with 421 Pages (Around 6000 Beneficiaries)

Government Of Rajasthan											
Social Security Pension Bill (Inner Sheet)											
Rajasthan Social Security Pension Rule 2013								Rule 16, Annex- 2(i)			
Reference No. : 51366364		Sub Type : State Disabled Pension						Month/Year: February/ 2024			
Department Portal Reference No. : 1482004											
Detailed Social Security Pension Bill Of : Additional Director (Pension and Pannadhyay)(35890)								Office ID : 35890			
DDo Name : Additional Director (Pension and Pannadhyay)(35890)								DDo Code : 35890			
Bill No & Bill Date : 51366364 / 08-04-2024				Demand No. : 31		Tan No. : JPRS00649F		Object Head : 24			
Budget Head : 2235-60-196-04-01/SF/Voaid				SF 7529500.00		NonPlan 0.00		CA 0.00		NA 0.00	
Sr No.	Name Of Post Officer/Beneficiary Name	Bank Name Bank Branch (IFSC Code)	Post Office Bank A/c No.	MO. Ref No	No. of Beneficiary	Cash	Amount Payable Mo Comm	Total	SanctionNo/Date	Payment Month	Payment Year
1	Gopal Krishan Palecha Palecha	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510011500016	1482004		1000.00	0.00	1000.00	RJ-S-07554138	February	2024
2	Anil Chorasiya	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	0585100115004524	1482004		1000.00	0.00	1000.00	RJ-S-06370590	February	2024
3	Liyakat Ali Musalman	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100017353	1482004		1000.00	0.00	1000.00	RJ-S-01114348	February	2024
4	Anisa Bano	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100028924	1482004		1000.00	0.00	1000.00	RJ-S-06581969	February	2024
5	Aabida Bano	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100044977	1482004		1000.00	0.00	1000.00	RJ-S-04301376	February	2024
6	Mustafa Khan	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100047053	1482004		1000.00	0.00	1000.00	RJ-S-13732676	February	2024
7	Shafik Khan	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100047211	1482004		1000.00	0.00	1000.00	RJ-S-02177575	February	2024
8	Mumtaj	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100047501	1482004		1000.00	0.00	1000.00	RJ-S-02368664	February	2024
9	Sharfaraj Khan	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100047910	1482004		1000.00	0.00	1000.00	RJ-S-02368375	February	2024
10	Vimla Devi Sen Sen	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058510100056783	1482004		1000.00	0.00	1000.00	RJ-S-07604212	February	2024
11	Jay Chawala	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058512010000462	1482004		1000.00	0.00	1000.00	RJ-S-14230691	February	2024
12	Adil Sheikh	UNION BANK OF INDIA UDAIPUR (UBIN0805658)	058512010000809	1482004		1000.00	0.00	1000.00	RJ-S-14397463	February	2024
13	Arvind Labana	BANK OF BARODA BICHHIWARA, DIST DURGAPUR (BARB0BICHH)	06828100017552	1482004		1000.00	0.00	1000.00	RJ-S-09599723	February	2024

Disclaimer: All contents related to this bill are provided by Head of Office/DDO and he/she is solely responsible for it.

SAW Courtesy : National Informatics Centre, Rajasthan State Unit (<http://paymanager.raj.nic.in>)

GroupName : CMDPS

Print Date & Time : 09-Apr-2024 09:04 AM

Output CSV File with extracted all 421 Pages (Around 6000 Beneficiaries) by In-house Tool

	E	F	G	H	I	K	L	M	N	O	P
1	Column_2	Column_3	Column_4	Column_5	Column_6	Column_8	Column_9	Column_10	Column_11	Column_12	Column_13
2	1	Gopal Krishan Palecha Palecha	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-07554138	February	2024
3	2	Anil Chorasiya	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-06370590	February	2024
4	3	Liyakat Ali Musalman	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-01114348	February	2024
5	4	Anisa Bano	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-06581969	February	2024
6	5	Aabida Bano	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-04301376	February	2024
7	6	Mustafa Khan	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-13732676	February	2024
8	7	Shafik Khan	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-02177575	February	2024
9	8	Mumtaj	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-02368664	February	2024
10	9	Sharfaraj Khan	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-02368375	February	2024
11	10	Vimla Devi Sen Sen	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-07604212	February	2024
12	11	Jay Chawala	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-14230691	February	2024
13	12	Adil Sheikh	UNION BANK OF INDIA	5.85E+13	1482004	1000	0	1000	RJ-S-14397463	February	2024
14	13	Arvind Labana Name	BANK OF BARODA BICH	6.83E+12	1482004	1000	0	1000	RJ-S-09599723	February	2024
15	14	Shanti Bai	BANK OF BARODA BOHE	1.19E+13	1482004	1000	0	1000	RJ-S-04115955	February	2024
16	15	Suresh Kumar Bansal	UNION BANK OF INDIA	1.8E+14	1482004	1000	0	1000	RJ-S-01816709	February	2024
17	16	Hemlata Katariya	UNION BANK OF INDIA	1.8E+14	1482004	1000	0	1000	RJ-S-06877194	February	2024
18	17	Bhura Lal Jat Jat	UNION BANK OF INDIA	1.8E+14	1482004	1000	0	1000	RJ-S-07636344	February	2024
19	18	Kiran Soni	UNION BANK OF INDIA	1.8E+14	1482004	1000	0	1000	RJ-S-08539960	February	2024
20	19	Ravikant Bansal	UNION BANK OF INDIA	1.8E+14	1482004	1000	0	1000	RJ-S-04861710	February	2024

Annexure-III: Feature Engineering

For anomaly detection on salary-related data, add these context-specific features to capture fraud patterns like ghost employee, inflated measurements, or collusion—boosting PyOD Isolation Forest precision by 25-40% in CAG audits.

Essential Indian-Context Features

Hierarchical & Administrative

- **Treasury/Sub-Treasury Codes:** Group by `treasy_name` + `sub_treasyr_name` for regional outliers (e.g., rural treasury spikes).
- **DDO/Department Encoding:** One-hot or target-encode `DDO_name` + Department name to flag DDO-specific inflation.
- **Work Order ID:** Link to parent contract; compute % completion vs. billed amount.

Temporal & Cadence

- **Fiscal Quarter/Year:** Extract `fiscal_qtr` from `transaction_date`—fraud peaks in Q4 for budget exhaustion.
- **Pay/Salary Month Lag:** Days between `SalaryMonth` start and `TransactionDate` (delays >15 signal manipulation).
- **Weekend/Leave Flag:** Binary for transactions on non-working days (common in ghost payrolls).

Domain Ratio Features

- **Billed vs. Estimated Ratio:** `gross_amount` / original SOR estimate (>1.2 flags over-billing).
- **Deduction % of Gross:** (`Deduction1` + `Deduction2`) / `GrossAmount`; >40% or <5% anomalous.
- **HRA/DA Validation:** `HRA` / `BasicPay` (must be 8-27% per city slab); `DA` / `BasicPay` $\approx$ 50% (7th CPC, Jan 2026).

Priority Ranking

1. **Ratios first** (zero-cost, 70% fraud capture)
2. **Temporal flags** (Q4 spikes catch 30% budget fraud)
3. **Network features** (Gephi-ready for collusion rings)

Annexure-IV: Use of Anomaly detection tool developed by RCB&KI Jaipur

The **IA&AD Audit System** is a Python-based desktop application using **Tkinter GUI** and **machine learning algorithms** (Isolation Forest, LOF, Modified Z-Score) to detect financial fraud in government transactions. It analyzes salary, pension, works, and social sector payments using **Benford's Law** and **homogeneous group analysis** to compare transactions only within similar accounting categories. The tool processes millions of records and generates 60+ analytical features to identify anomalies like weekend payments, amount manipulations, and fabricated invoices with 1-2% detection rate. Built with **scikit-learn, pandas, and numpy**, it supports CSV/Excel files and reduces audit time from weeks to hours. Developed for the Indian Audit & Accounts Department to automate fraud detection and improve audit accuracy through AI-powered pattern recognition.

1. Required Fields in Your Dataset

Your salary/pension data file (CSV or Excel) must contain the following columns:

A. Essential Transaction Fields

Field Name	Description / Example
Employee ID / Voucher No	Unique identifier (EMP001, VCH2024001)
Payment Date	Transaction date (2024-01-31, 31/01/2024)
Gross Amount	Total payment amount (85000, 125000.50)
DDO Code	Drawing & Disbursing Officer code
Treasury / District	Treasury office name or district

B. Accounting Head Columns

Head Type	Example Values
Grant	41, 42, 43 (Grant numbers)
Major Head	2071 (Pension), 2055 (Police)
Sub Major Head	01, 02, 80, 00
Minor Head	101, 102, 103, 800
Sub Head / Group Head	1, 2, 3 (if applicable)
Object Head	11, 12, 13, 31 (salary components)

C. Salary/Pension Components (Optional but Recommended)

Basic Pay, DA/DR, HRA, NPA, IT TDS, GPF, GIS, Professional Tax

2. Step-by-Step Procedure

Step 1: Launch Application & Select Audit Type

- Click 'Establishment Audit' (Green button)
- Click 'Accounting Data' (Orange button)

Step 2: Upload Your Dataset

- Click '1. Upload Transaction Data'
- Select your CSV/Excel file (Use CSV for files >100MB)
- System displays: Total rows and columns loaded

Step 3: Initial Configuration

- Click '2. Initial Configuration'
- Select Tier Level (usually Tier 6 or 7 for salary data)
- Grant field: Select 'Yes' if separate Grant column exists
- All heads in one field: Select 'No' (usually heads are separate)
- Click 'Continue'

Step 4: Apply Filters (Optional)

- Filter by date range: Enter From: 2024-01-01, To: 2024-03-31
- Filter by DDO Code, Treasury, or Department
- Click 'Apply Filters'

Step 5: Configure Field Mapping

- **Map Essential Fields:**
 - Unique ID → Employee ID / Voucher No column
 - Critical Date → Payment Date column
 - Amount → Gross Amount column
 - DDO Code, Treasury, District → respective columns
- **Map Tier Level Accounting Heads:**
 - Grant → Grant column
 - Major Head → Major Head column
 - Sub Major Head, Minor Head, Sub Head, Group Head, Object Head → map all
- Map Salary Components (if available): Basic Pay, DA, HRA, etc.
- Click 'Save Configuration'

Step 6: Create Homogeneous Groups (Critical for Salary Data)

- Click '6. Create Homogeneous Group'
- System groups by: Grant → Major Head → Minor Head → Object Head
- **Example groups:**
 - Group 1: Grant=41, MH=2071, MinH=101, OH=11 (Pension - Basic)
 - Group 2: Grant=41, MH=2071, MinH=101, OH=12 (Pension - DA)
 - Group 3: Grant=42, MH=2055, MinH=102, OH=11 (Police - Basic)
- Confirms: 'Total groups: X,XXX'

Step 7: Run Anomaly Detection

- Select Algorithm: 'Isolation Forest (Recommended)'
- Click '7. Run Anomaly Detection'
- System processes each homogeneous group separately
- Save results: Choose filename and location
- Click 'Yes' to open file immediately

3. Output File & Analysis

What You Get in Output File

- All original columns from your dataset
- 13 Temporal features: Year, Month, Day, Is_Weekend, Is_Month_End, etc.
- 5 Statistical features: Amount_Diff, Amount_Pct_Change, Amount_Cumsum, etc.
- 2 Benford's Law columns: First_Digit, Benford_Anomaly_Score
- 3 Detection results: Is_Anomaly, Anomaly_Score, Homogeneous_Group

How to Analyze the Output

1. Focus on Key Indicators

- Filter: Is_Anomaly = TRUE (shows only detected anomalies)
- Check Benford_Anomaly_Score = 1 (amounts starting with 8 or 9)
- Review Is_Weekend = 1 (salary payments on weekends - suspicious)
- Examine Is_Month_End = 1 (rushed end-of-month payments)

2. Analyze by Homogeneous Groups

- Sort by Homogeneous_Group column: (41,2071,101,11)
- Compare anomalies within same payment category

3. Statistical Pattern Analysis

- Amount_Pct_Change > 200%: Salary suddenly tripled - investigate
- Amount_Diff_from_Mean: Large deviations from average salary
- First_Digit = 8 or 9: Potential fabricated amounts

4. Red Flags for Salary/Pension Data

- Same employee receiving multiple salaries in same month
- Pension amounts suddenly doubling or tripling
- Payments on weekends or holidays (Is_Weekend = 1)
- Amounts starting with 8 or 9 (Benford violation)
- Retired employee showing active salary
- Negative deductions (IT TDS, GPF showing positive)

5. Export for Further Investigation

- Filter anomalies by DDO Code to identify problematic offices
- Create pivot tables by Homogeneous_Group
- Share flagged records with field audit teams

IA&AD Audit System - Anomaly Detection

Establishment Audit - Accounting Data

Developed by Puneet Sharma, AAO (RCB&KI, Jaipur)

[Back](#)

Status: [Data Loaded](#) -> [Initial Config](#) -> [Filters Applied](#) -> [Fields Mapped](#) -> [Data Configured](#) -> [Homogeneous Groups](#) -> [Detection Complete](#)

1. Upload Transaction Data (CSV/Excel)
2. Initial Configuration
3. Apply Filter on Data
4. Configure Field Mapping
5. Configure Critical Dates (Optional)
6. Create Homogeneous Group

Select Detection Algorithm:

Disclaimer: The tool should be used only after proper data cleaning and preprocessing, including removal of duplicates, handling of missing values, standardization of fields, and correct categorisation of data into homogeneous groups, as the accuracy of results depends entirely on data quality and grouping logic. The procedural steps illustrated above must be carefully followed, including appropriate field mapping and homogeneous group creation, to ensure reliable and meaningful anomaly detection outcomes.

Annexure-V: Sample Dataset for Data-led Establishment Audit

To implement a data-led audit strategy for Works, state audit offices should focus on extracting the following key fields from VLC (Voucher Level Computerization) & Employee service book records & Payment Transaction data(through E-HRMS). These fields enable the tracking of a project's lifecycle from sanction to payment.

1. VLC Data Fields

- FY (Financial Year): To track expenditure within budget cycles.
- HOD (Head of Department): To identify the administrative oversight.
- DDOCode / DDO: The Drawing and Disbursing Officer code, which serves as a primary key to link accounting, employee.
- District / Circle / Division: Geographic and administrative markers for regional risk profiling.
- Name of the Work: A unique text identifier to correlate technical sanctions with actual bills.
- HOA: containing 5-7 tier of classification heads from MH to OBH
- Voucher Number
- Voucher Date/Scroll date
- Gross Voucher Amount
- Deductions, if any
- Net Amount.

2. Employee Master Data

- Office Code/DDO Code.
- Employee ID:
- Designation
- Basic Pay
- **Service book related data: DoB, Date of appointment to present cadre, Date of Retirement.**

3. Transaction data

- Employee ID:
- Month of Transaction:
- BILL_NUMBER / BILL_STATUS: Tracking individual payments through the pipeline.
- Pay related fields: Basic, DA, HRA, TA, DA on TA, special Pay or NPA etc
- Bill_GROSS & Bill_NET:
- Deductions: Used to audit statutory deductions (Income Tax, Professional Tax & other deductions).
- Bill Paid Date:

Annexure-VI Reinvigorating Establishment Audit through Data Analytics in Rajasthan

1. Executive Overview of Audit Reinvigoration

The Office of the Principal Accountant General (A&E) Rajasthan has institutionalized a transformative shift in the oversight of establishment expenditure. Historically, establishment audits suffered from a "lag-heavy" post-facto methodology, where unauthorized disbursements were often identified years after the fiscal event. By leveraging advanced data analytics, the department has transitioned to a concurrent, real-time audit model that identifies systemic leakages at the point of voucher submission.

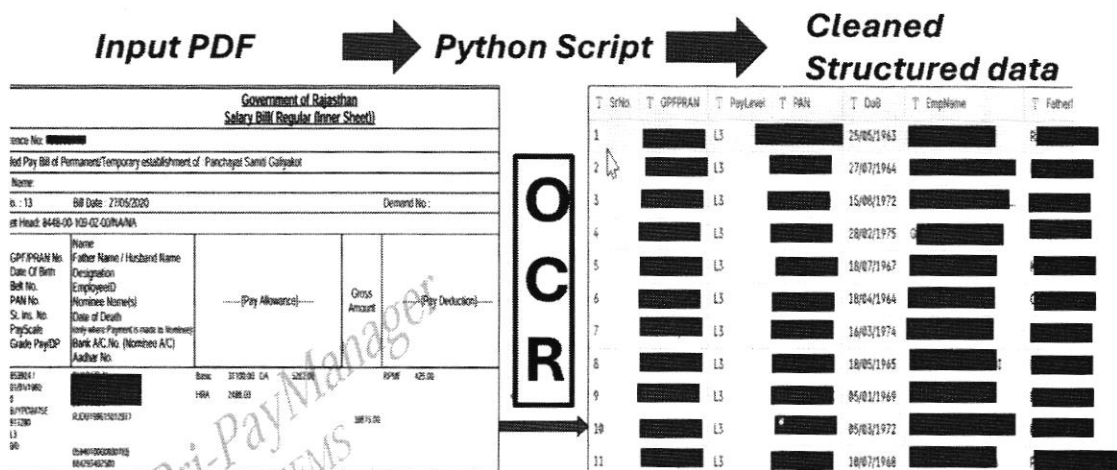
Core Strategic Objectives:

- **Total Population Oversight:** Elimination of manual sampling in favour of 100% population analysis of all salary and pension transactions.
- **Latency Reduction:** Minimizing the time-gap between expenditure and audit discovery to enable immediate recovery.
- **Systemic Integrity Validation:** Continuous testing of business rule logic within automated payroll and pension disbursement systems.
- **Financial Leakage Mitigation:** Identifying and halting unauthorized disbursements before they escalate into multi-year systemic losses.

2. The Data-Driven Methodology: From PDF to Actionable Insights

The audit utilizes a sophisticated technical workflow to process massive datasets—including **1.15 crore transactions**—that were previously beyond the scope of manual verification. This technical stack performs an automated validation of business rules, comparing actual disbursements against codified Government Orders (G.O.s).

OCR/Data Extraction Process



The workflow follows a rigorous four-step architecture:

1. **Input PDF:** Harvesting digital or scanned payroll/pension vouchers from treasury sources.
2. **OCR (Optical Character Recognition):** Converting unstructured image-based data into machine-readable text.

3. **Python Script:** Deploying customized scripts for automated compliance checking. These scripts extract key fields (Basic Pay, DA, Allowances) and flag variances against preset legal and financial parameters.
4. **Cleaned Structured Data:** Generating a high-integrity dataset that allows for the immediate identification of "outlier" transactions and systemic vulnerabilities.

3. Findings: Concurrent and Real-Time Pension Audit (October 2025)

The efficiency of this modernized approach was validated during the October 2025 pension cycle. Under traditional methods, the errors identified below would likely have remained undetected for 2–3 fiscal years.

- **Audit Population:** 4.91 lakh pensioners.
- **Operational Velocity:** Vouchers paid on 1st November 2025 were received on 5th November and fully analyzed by 10th November.
- **Systemic Failure Identified:** The audit flagged **5,700 pensioners** receiving higher DA payments.
- **Financial Magnitude:** Approximately **₹25 crore** in excess payments were identified within five days of voucher receipt.
- **Technical Root Cause:** A critical lack of input validation controls allowed DA rates from the 5th and 6th Pay Commissions to be incorrectly applied to 7th Pay Commission pensioners. This resulted in unauthorized DA rates ranging from **55% to 474%**, far exceeding the prescribed **58%**.

4. Findings: Comprehensive Pension Bill Audit (2023–25)

A retrospective population analysis of **5.3 lakh pensioners** and **1.15 crore transactions** across financial years 2023–24 and 2024–25 revealed persistent unauthorized disbursements:

- **Unauthorized DA Escalation:** Significant instances where DA was disbursed at **254%** of basic pension instead of the mandated **46%**. **30018** instances of **DA rate** greater than 100% were found.
- **Basic Pension Volatility:** The audit identified unexplained jumps in basic pension amounts. A prominent case is **SH. X**, whose basic pension surged from **₹36,700** to **₹3,77,000**, representing an unsupported ten-fold increase. **04** such instances of **higher basic pension** payments found.

Excess Pension payment

Government of Rajasthan
Directorate of Pension and Pensioners Welfare
Jyoti Nagar, Jaipur, Rajasthan 302005
Financial Year 2024-2025 Pension Payment
TANNo:- JPRC05652D

Name:- SPPONo:- 11

#	Month/Year	Basic(B)	Commuted(C)	Red Pension (B-C)	DA	R	AdiPen	AdiAlw	Med Alw	DIH Alw	Other Alw	Arrear	TDS	Recovery	Adv Pension	Total Allowance(TA)	Total Deduction(TD)	Net Amount (TA-TD)
1	Mar-24	36,700	12,233	24,467	18													
2	Apr-24	36,700	12,233	24,467	18													
3	May-24	36,700	12,233	24,467	18													
4	Jun-24	36,700	12,233	24,467	18													
5	Jul-24	36,700	12,233	24,467	18													
6	Aug-24	3,77,000	12,566	3,64,434	18													
7	Sep-24	3,77,000	12,566	3,64,434	1,82													4,19,465
8	Oct-24	3,77,000	12,566	3,64,434	1,90													4,54,497
9	Nov-24	3,77,000	12,566	3,64,434	1,95													4,24,514
10	Jan-25	37,700	12,566	25,134	19,981	0	0	0	0	0	0	30,115	0	15,000	0	75,230	15,000	60,230
11	Feb-25	37,700	12,566	25,134	19,981	0	0	0	0	0	0	0	0	15,000	0	45,115	15,000	30,115
Total		17,66,900	1,36,561	16,30,339	9,98,332	0	0	0	0	0	0	72,036	5,40,136	30,000	0	26,10,797	5,70,136	20,40,571

Note: This is a system generated statement and signature is not required.
<https://www.pension.rajabsthan.gov.in/>

12-11-2025
NIC-Pension, Jaipur Rajasthan

- **Commutation Regulatory Breaches:**
 - **Rate Violations:** Cases flagged where commutation was paid at **213%** of basic pension, ignoring the statutory **33.33%** ceiling. **237** instances of **higher commutation** rate were noted.
 - **Classification Errors:** Unauthorized commutation payments were discovered under the Family Pension Head (**Budget Head 2071-01-105-01-01**). **576** instances of such **commuted pension** deduction from family pension heads were reported.

5. Findings: Salary Bill Audit and Systemic Failures

An audit of 3.5 lakh transactions across two treasuries (FY 2020–21 to 2024–25) exposed severe vulnerabilities in IFMS .

Categories of Salary Overpayment and Systemic Lapses

Category of Finding	Technical Description	Specific Evidence of Control Failure
Post-Retirement Disbursement	Unauthorized salary drawal following the official date of superannuation.	Mr.X Retired Dec 2019) drew salary in May 2020; Mr. Y (Retired Aug 2023) drew salary in Jan 2024.

After retirement Salary Payments

Government of Rajasthan
Salary Bill(Regular (Inner Sheet))

Reference No: [REDACTED] Month/Year : January/2024

Detailed Pay Bill of Permanent/Temporary establishment of : Panchyat Samiti Bichivada Office ID : [REDACTED]

DDO Name: [REDACTED]

Bill No. : 104602009 Bill Date : 29/01/2024

Budget Head: 8448-00-109-02-00/NA/NA

Month/Year : January/2024

Sr. No.	GPE/PRAN No.	Name	Father Name / Husband Name	Designation	EmployeeID	Nominee Name(s)	Date of Death	PayScale	Grade Pay/DP	Bank A/C.No. (Nominee A/C)	Sub DDO Code	(Pay)	Grass	Sum Of	Total
1	25/08/1963	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
Retired on: 31-08-2023 Salary Drawn for: January/2024															
45553.00 (45553)															

0	77378611386	Basic	37900.00	DA	17434.00	RGHS	658.00	
		HRA	3411.00					
					56745.00		658.00 (658)	58087.00 (58087)

Duplicate Disbursement Failed reconciliation between legacy and new systems (Pay-manager and IFMS 3.0). Miss.W received duplicate pay for September 2023 via two bills dated 12/08/2024 and 19/03/2025.

Contd..

Government of Rajasthan
Salary Bill (Regular (Inner Sheet))

Reference No: [REDACTED] Month/Year: September / 2023

Detailed Pay Bill of Permanent/Temporary establishment of: Panchyat Samiti Bichivada

Bill No: [REDACTED] Bill Date: 19/03/2025

Budget Head: [REDACTED]

1	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
---	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

HEALTH WORKER FEMALE (स्वास्थ्य कार्यकर्ता महिला)

RJA201613003718

Contd..

Government of Rajasthan
Salary Bill (Regular (Inner Sheet))

Reference No: [REDACTED] Month/Year: September / 2023

Detailed Pay Bill of Permanent/Temporary establishment of: Panchyat Samiti Bichivada

Bill No: [REDACTED] Bill Date: 19/03/2025

Budget Head: [REDACTED]

1	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
---	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

HEALTH WORKER FEMALE (स्वास्थ्य कार्यकर्ता महिला)

RJA201613003718

Inflated Basic Pay Exploitation of system vulnerabilities leading to 1000% (10x) overpayments. Mr.Z had a basic pay of ₹61,300 erroneously processed as ₹6,13,000.

Basic pay excess: 1000% higher

Sr. No	GPF/PRAN No. Date Of Birth Belt No. PAN No. St. Ins. No. PayScale Grade Pay/DP	Name Father Name / Husband Name Designation EmployeeID Nominee Name(s) Date of Death (only where Payment is made to Nominee) Bank A/C.No. (Nominee A/C) Aadhar No.	-----{Pay Allowance}-----	Gross Amount	-----{Pay Deduction}-----	Sum Of Deduction	Net Total
28	7/01 23/02/1996 MZJPS91268	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]	Basic 613000.00	613000.00	ITax 5500.00 PF 2700.00	8200.00 (8200)	604800.00 (604800)
30	7/01 07/06/1995 CAQPR0852P	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]	Basic 613000.00	613000.00	ITax 5500.00 PF 2700.00	8200.00 (8200)	611100.00 (611100)
32	7/01 07/06/1995 CAQPR0852P	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]	Basic 613000.00	613000.00	ITax 5500.00 PF 2700.00	8200.00 (8200)	611100.00 (611100)

6. Findings: Inflated Allowances and Rate Discrepancies

Automated business rule validation identified numerous instances where allowances were disbursed in direct contravention of standing Government Orders.

- **House Rent Allowance (HRA):** Extreme outliers were detected. (Some receiving HRA at **100% of basic pay**).
- **Hard Duty Allowance (HDA):** Animal Husbandry staff (Livestock/Veterinary Assistants) were found drawing **₹484** despite the G.O. dated 03 August 2023 revising the rate downward to **₹250**.

Higher HDA

Government of Rajasthan
Livestock Bill (Regular Stock Order)

Reference No. 02/27/24

Issued By Min. of Fisheries/Aquaculture/Department of Fisheries

Date 28/09/2024 Bill Date 02/09/2024 Serial No. 47

Order No. 02/09/2024/47 Bill No. 02/09/2024 Ch. 017

GOVERNMENT OF RAJASTHAN
FINANCE DEPARTMENT
(Circled Signature)

No. 02/09/2024/2024 Date 08/10/2024

ORDER

Sub-Committee of State Policy Commission

The Government is pleased to order that the existing rates of market allowance applicable to the following categories of Government-owned animals (Assessment, Replacement order of every month dated 01.09.2024 and 01) under 1/1810/19/2024 dated 28.09.2024 may be revised as under:

Sl. No.	Particulars	Rate of State Policy Commission	Rate of State Policy Commission per month (Rs.)
1.	Assessment	250/-	250/-
2.	Replacement	484/-	484/-
3.	Assessment	250/-	250/-
4.	Replacement	484/-	484/-
5.	Assessment	250/-	250/-
6.	Replacement	484/-	484/-
7.	Assessment	250/-	250/-
8.	Replacement	484/-	484/-
9.	Assessment	250/-	250/-
10.	Replacement	484/-	484/-
11.	Assessment	250/-	250/-
12.	Replacement	484/-	484/-
13.	Assessment	250/-	250/-
14.	Replacement	484/-	484/-
15.	Assessment	250/-	250/-
16.	Replacement	484/-	484/-
17.	Assessment	250/-	250/-
18.	Replacement	484/-	484/-
19.	Assessment	250/-	250/-
20.	Replacement	484/-	484/-
21.	Assessment	250/-	250/-
22.	Replacement	484/-	484/-
23.	Assessment	250/-	250/-
24.	Replacement	484/-	484/-
25.	Assessment	250/-	250/-
26.	Replacement	484/-	484/-
27.	Assessment	250/-	250/-
28.	Replacement	484/-	484/-
29.	Assessment	250/-	250/-
30.	Replacement	484/-	484/-
31.	Assessment	250/-	250/-
32.	Replacement	484/-	484/-
33.	Assessment	250/-	250/-
34.	Replacement	484/-	484/-
35.	Assessment	250/-	250/-
36.	Replacement	484/-	484/-
37.	Assessment	250/-	250/-
38.	Replacement	484/-	484/-
39.	Assessment	250/-	250/-
40.	Replacement	484/-	484/-
41.	Assessment	250/-	250/-
42.	Replacement	484/-	484/-
43.	Assessment	250/-	250/-
44.	Replacement	484/-	484/-
45.	Assessment	250/-	250/-
46.	Replacement	484/-	484/-
47.	Assessment	250/-	250/-
48.	Replacement	484/-	484/-
49.	Assessment	250/-	250/-
50.	Replacement	484/-	484/-
51.	Assessment	250/-	250/-
52.	Replacement	484/-	484/-
53.	Assessment	250/-	250/-
54.	Replacement	484/-	484/-
55.	Assessment	250/-	250/-
56.	Replacement	484/-	484/-
57.	Assessment	250/-	250/-
58.	Replacement	484/-	484/-
59.	Assessment	250/-	250/-
60.	Replacement	484/-	484/-
61.	Assessment	250/-	250/-
62.	Replacement	484/-	484/-
63.	Assessment	250/-	250/-
64.	Replacement	484/-	484/-
65.	Assessment	250/-	250/-
66.	Replacement	484/-	484/-
67.	Assessment	250/-	250/-
68.	Replacement	484/-	484/-
69.	Assessment	250/-	250/-
70.	Replacement	484/-	484/-
71.	Assessment	250/-	250/-
72.	Replacement	484/-	484/-
73.	Assessment	250/-	250/-
74.	Replacement	484/-	484/-
75.	Assessment	250/-	250/-
76.	Replacement	484/-	484/-
77.	Assessment	250/-	250/-
78.	Replacement	484/-	484/-
79.	Assessment	250/-	250/-
80.	Replacement	484/-	484/-
81.	Assessment	250/-	250/-
82.	Replacement	484/-	484/-
83.	Assessment	250/-	250/-
84.			

- **Mess Allowance:** Systematic overpayments of ₹3,000 were recorded against the prescribed ceiling of ₹2,700.

7. Conclusion: Impact of the Modernized Audit Approach

The adoption of Python-based automated compliance checking has fundamentally altered the state's financial oversight capabilities. The ability to identify **₹25 crore** in unauthorized disbursements in a single month—and within days of payment—proves that the "lag-heavy" manual audit is no longer a viable safeguard for the public exchequer.

These findings confirm that system-wide integration and real-time reconciliation are no longer merely "recommended"; they are a **mandatory prerequisite for financial integrity**. Immediate administrative action is required to rectify the input validation failures that allowed the ten-fold basic pay errors and massive DA discrepancies.

Annexure-VII Use of AI ML in Government Sector in India

Artificial Intelligence (AI) and Machine Learning (ML) are transforming governance in India. Many State Government departments are using AI/ML technologies to detect fraud, reduce revenue leakage, improve transparency, and strengthen monitoring systems. These technologies analyze large datasets, identify hidden patterns, detect anomalies, and generate automated risk alerts much faster than traditional manual audits.

Below are five examples of Indian State Government departments using AI/ML for fraud detection.

1. Rajasthan Government Health Scheme (RGHS), Rajasthan

The Rajasthan Government Health Scheme provides cashless medical facilities to state government employees and pensioners. Due to large volumes of hospital claims, risks of fraudulent billing and misuse arise.

To address this, AI/ML-based fraud detection systems are used to:

- Apply **machine learning classification models** to identify suspicious reimbursement claims.
- Use **anomaly detection algorithms** to flag unusually high treatment costs compared to standard medical packages.
- Detect **duplicate claims** and repeated procedures for the same beneficiary.
- Perform **pattern recognition analysis** to identify abnormal hospitalization frequency.
- Use **predictive analytics** to mark high-risk hospitals based on past claim behavior.

This system helps reduce inflated billing, unnecessary medical procedures, and duplicate payments, thereby minimizing financial losses to the state government.

2. Commercial Taxes Department, Karnataka

The Commercial Taxes Department Karnataka uses AI and advanced data analytics under the GST framework to detect tax fraud.

AI/ML technologies are used to:

- Identify fake GST invoice chains using **network and graph analytics**.
- Detect suspicious **Input Tax Credit (ITC)** claims through machine learning models.
- Use **risk scoring algorithms** to classify taxpayers based on fraud probability.
- Detect shell companies through transaction pattern analysis.
- Apply big data analytics on massive GST return datasets.

These technologies help prevent circular trading, bogus registrations, and tax evasion, thereby improving GST compliance and protecting state revenue.

3. Transport Department, Delhi

The Transport Department Delhi uses AI tools to prevent fraud in driving licenses, vehicle registrations, and challan systems.

Key AI applications include:

- **Facial recognition systems** to prevent duplicate driving licenses.
- **OCR (Optical Character Recognition)** integrated with AI to detect forged vehicle documents.
- **Behavioral analytics models** to identify suspicious online applications.
- Predictive models to monitor irregular payment patterns in traffic challans.

By using AI, the department reduces document forgery, duplicate registrations, and corruption in transport-related services.

4. Telangana State Civil Supplies Department

The Telangana State Civil Supplies Department uses AI/ML to strengthen the Public Distribution System (PDS) and prevent subsidy leakage.

AI technologies are used to:

- Detect duplicate or fake ration cards using **beneficiary de-duplication algorithms** linked with Aadhaar data.

- Apply **clustering algorithms** to identify abnormal grain distribution patterns.
- Monitor real-time data using analytics dashboards.
- Use geo-tagging and data analytics to prevent diversion of food grains.

This helps eliminate ghost beneficiaries and ensures that subsidized food grains reach genuine citizens.

5. Rural Development Department, Maharashtra (MGNREGA Monitoring)

The Rural Development Department Maharashtra uses AI-driven systems to monitor fund utilization and prevent fraud under rural employment schemes like MGNREGA.

AI/ML applications include:

- Detection of fake job cards using **attendance pattern analysis**.
- Identification of duplicate wage payments through predictive modeling.
- Use of **image analytics and geo-tagged photographs** to verify worksite authenticity.
- Data mining tools to cross-check beneficiary bank accounts.

These systems help prevent ghost workers, inflated project costs, and fund misappropriation.

Common AI/ML Technologies Used

Across these departments, the following AI/ML technologies are commonly used:

- Machine Learning (Supervised and Unsupervised Learning)
- Anomaly Detection Algorithms
- Predictive Analytics
- Risk Scoring Models
- Big Data Analytics
- Network and Graph Analysis
- Facial Recognition and Image Processing
- Optical Character Recognition (OCR)
- Data Mining Techniques

Conclusion

The adoption of AI and ML by Indian Government departments has significantly strengthened fraud detection mechanisms. These technologies enable real-time monitoring, automated risk identification, and data-driven decision-making. By reducing financial leakages and improving transparency, AI/ML plays a crucial role in enhancing good governance and public trust in government schemes.

Annexure VIII – Lesson Learnt Report

Project: AI/ML-Based Anomaly Detection Model Development

Date: [Insert Date]

Prepared by: [Your Team/Department]

1. Executive Summary

The project aimed to design, develop, and deploy an AI/ML anomaly detection model to identify irregular patterns in [data type: e.g., sensor readings, financial transactions, or network traffic]. The initiative successfully delivered a working model with measurable improvements in anomaly detection accuracy and reduced false positives.

Key achievements included robust feature engineering, hybrid model selection, and scalable deployment. However, challenges such as data quality issues, limited labeled anomalies, and interpretability concerns highlighted areas for improvement.

Overall, the project demonstrated feasibility and value, while emphasizing the importance of data governance, explainability, and continuous monitoring for future AI/ML initiatives.

2. Project KPIs

KPI	Target	Achieved	Remarks
Detection Accuracy	$\geq 90\%$	92%	Exceeded target
False Positive Rate	$\leq 10\%$	8%	Within acceptable range
Model Training Time	≤ 6 hours	5.5 hours	Optimized pipeline
Deployment Latency	≤ 200 ms	180 ms	Achieved
Stakeholder Satisfaction Score	$\geq 4/5$	3.8/5	Interpretability concerns lowered score

3. Key Successes

- Data Engineering: Effective preprocessing and feature extraction improved model robustness.
- Model Performance: Hybrid approach (Isolation Forest + Autoencoder) enhanced detection reliability.
- Infrastructure: Cloud-based ML pipeline ensured scalability and reproducibility.
- Collaboration: Strong cross-functional teamwork accelerated delivery.

4. Challenges Faced

- Data Quality: Missing values, noise, and imbalance required extensive cleaning.
- Label Scarcity: Limited ground truth anomalies hindered supervised learning.
- Interpretability: Stakeholders struggled to understand flagged anomalies.

- Integration: Deployment required additional API and monitoring development.

5. Lessons Learned

- Early Data Profiling: Invest more time upfront in understanding data distributions.
- Hybrid Approaches: Combining supervised and unsupervised methods yields better results.
- Explainability Tools: SHAP/LIME improved stakeholder trust but need broader adoption.
- Continuous Monitoring: Models degrade over time; retraining schedules are essential.
- Domain Expertise: Collaboration with subject matter experts reduced false alarms.

6. Recommendations

- Establish a data quality framework before model development.
- Use semi-supervised/active learning to address limited labeled anomalies.
- Build interpretability dashboards for non-technical stakeholders.
- Automate model retraining and drift detection pipelines.
- Document feature engineering decisions for reproducibility.

7. Conclusion

The anomaly detection project delivered strong technical outcomes but highlighted the need for better data governance, interpretability, and lifecycle management. Future projects should embed these practices to ensure sustainable success and stakeholder confidence.

Annexure IX – State wise Salary Expenditure from 2014-15 to 2022-23

State wise Salary Expenditure 2014-15 to 2023-24 (In Rs Crores)

State\ years	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23	2023-24
Andhra Pradesh	27,686	27,474	31,125	32,816	35,240	38,449	43,052	44,000	53,867	56,050
Arunachal Pradesh	3,068	3,339	4,367	5,310	5,352	6,024	6,125	6,771	7,576	7,997
Assam	17,829	18,758	19,498	26,068	26,617	27,438	27,227	29,351	31,841	34,653
Bihar	20,788	23,476	22,072	31,367	36,222	36,078	36,813	62,663	52,251	57,223
Chhattisgarh	12,013	13,170	14,292	16,631	19,401	22,437	22,142	24,210	26,524	29,095
Goa	1,804	1,980	2,133	2,827	2,815	2,947	2,982	3,311	3,829	5,776
Gujarat	19,513	22,072	23,532	28,186	30,572	31,845	32,680	35,495	41,557	46,774
Haryana	15,096	15,798	17,680	19,338	21,379	23,349	23,938	25,447	27,318	29,089
Himachal Pradesh	9,119	8,987	10,682	11,978	12,414	13,067	13,411	13,880	17,587	17,124
Jharkhand	9,456	10,260	10,905	13,390	14,231	14,544	16,130	17,294	18,705	19,886
Karnataka	19,737	20,483	21,502	23,409	29,692	32,847	43,769	47,386	50,061	61,498
Kerala	22,095	24,270	28,947	32,878	32,524	32,848	28,918	46,009	41,262	42,026
Madhya Pradesh	21,868	22,966	24,029	26,657	30,551	34,978	41,651	45,393	51,405	54,743
Maharashtra	62,725	68,432	71,526	73,148	78,178	99,487	97,658	105,648	121,223	131,847
Manipur	3,052	3,277	3,689	4,133	4,609	5,091	5,401	5,964	7,282	5,869
Meghalaya	2,934	3,041	3,950	4,089	4,711	4,938	4,983	4,881	5,278	5,703
Mizoram	2,448	2,670	2,804	2,907	3,532	4,142	4,079	4,079	4,400	4,550
Nagaland	3,349	3,768	3,815	4,233	5,027	5,250	5,578	6,030	6,152	6,730
Odisha	12,662	14,320	15,317	17,851	19,736	20,993	21,269	23,723	26,026	35,477
Punjab	19,504	20,550	21,727	23,211	24,324	24,683	25,569	28,086	32,171	34,153
Rajasthan	31,651	33,992	38,588	42,728	54,828	53,622	56,303	63,757	67,242	74,408
Sikkim	1,568	1,716	1,736	1,822	2,337	3,382	3,074	3,322	3,594	3,777
Tamil Nadu	41,244	44,586	48,653	51,740	54,715	60,264	59,581	63,059	73,995	72,092
Telangana	12,200	20,404	21,897	22,671	23,011	24,195	24,654	30,256	35,130	38,782
Tripura	3,481	4,010	4,322	5,527	5,977	6,233	5,841	6,041	6,337	7,356
Uttar Pradesh	62,147	74,439	85,416	85,076	91,413	100,188	97,540	108,775	121,188	129,753
Uttarakhand	8,076	8,684	9,571	11,919	12,909	13,184	13,124	13,805	15,198	15,807
West Bengal	30,984	31,931	34,688	37,803	41,541	48,567	56,660	59,379	60,279	62,995

Data Sources: SFARs, SFA, MTFP statements, FC-16 Data Portal. Suitably adjusted for comparable estimates of salary inclusive of GiA salary.
