**1.Step-by-Step Summary of the Code Workflow**

**1. Custom Embedding with LM Studio**

* A local embedding model (via LM Studio at http://localhost:1234/v1/embeddings) is used to transform resume/job content into vector representations.
* LMStudioEmbeddings class handles:
  + embed\_documents() – converts resume chunks into embeddings for indexing.
  + embed\_query() – converts JD into embedding to search relevant resume chunks.

**2. File Readers**

* Resume and JD are extracted using:
  + read\_pdf\_text() – for PDFs (e.g., resumes).
  + read\_docx\_text() – for DOCX files (e.g., JDs).
  + read\_txt\_file() – for reference samples or manual versions.

**3. Scrubbing PII**

* Applies regex to remove:
  + Phone numbers, emails, URLs, company names, cities, and years.
  + Ensures anonymized and clean data for consistent LLM behavior.

**4. Resume Chunk Extraction & Matching**

* Sentences split and ranked using:
  + TF-IDF + Cosine Similarity to find top-k lines most similar to JD.
  + Appends bottom part of resume (last 2000–3000 chars) to retain senior roles.
  + Optionally tagged into categories: "Model Validation" vs "Other Experience".

**5. Knowledge Ingestion for RAG (Retrieval-Augmented Generation)**

* Web documents (Basel, EBA, BoE) are scraped and stored in FAISS index using LM Studio embeddings.
* build\_credit\_faiss\_store() – splits, embeds, indexes documents.
* build\_rag\_context\_from\_credit\_knowledge() – retrieves top 6 chunks relevant to JD.

**6. Prompt Construction**

* Combines:
  + Extracted resume content
  + JD
  + RAG context
  + EY format reference sample
* Builds a structured prompt aligned to EY expectations in tone, format, and keyword density.

**7. LLM Call (via Mistral model)**

* Sends prompt to local LLM endpoint /v1/completions hosted by LM Studio.
* Returns customized resume aligned to EY’s expectations.

**8. Post-Processing / Output Patching**

* patch\_llm\_output():
  + Ensures keywords like “SR 11-7”, “RNIV”, “Z-spread”, “XGBoost” are present.
  + Adds model types (PD, LGD, EAD) if missed.
  + Pulls missing terms from EY reference resumes.

**9. Similarity Scoring**

* Computes cosine similarity using TF-IDF between:
  + LLM-generated output and EY Sample → Format Compliance Score
  + LLM-generated output and Manual Version → Content Personalization Score

|  |  |
| --- | --- |
| **Manual Match Score (e.g., 61–66%)** | **Measures how close the content is to a human-customized resume (personalization and examples).** |
| **Score Drop (e.g., from 65% → 63%)** | **Attributed to reduced top-k content (from 20 to 8 lines), patching interference, phrasing gaps.** |

**2.LLMs Learn Better with Examples**

* The LLM’s output quality significantly improves when it sees **diverse sample resumes across formats, domains, and geographies.**
* Current samples may be **biased toward one style or structure, reducing generalizability.**

**2. Limited Samples = Poor Contextual Learning**

* Without enough high-quality examples, the model:
  + **Misses nuanced phrasing used by domain experts.**
  + **Fails to generalize across roles (validation, monitoring, IFRS 9, climate risk).**
  + **Produces repetitive or generic outputs.**

**3. Few-Shot Prompting Depends on Reference Diversity**

* Few-shot prompting requires strong anchor samples.
* With just 1–2 resumes and 1 JD, the model’s response range is narrow.
* Adding 10+ diverse synthetic resumes (from other candidates or generated) helps the model learn *how to adapt structure across cases*.

4. Improved RAG Context for Better Prompts

* RAG-augmented prompts pull in more domain-rich content when more sample resumes are embedded.
* This enhances context retrieval → improves accuracy, coverage, and compliance.

5. Boosts Match Score to Manual Versions

* Every 5–10 enriched synthetic samples increases the manual match score by:
  + Teaching the model industry-specific phrasing.
  + Reinforcing missing keyword placement.
  + Giving better prompts through FAISS-retrieved context.

| **3.Traditional Python Logic** |
| --- |
| Rigid keyword match; no nuance |
| Cannot personalize beyond pre-coded templates |
| Cannot infer—only matches explicit presence |
| Requires rigid, preprocessed input |
| Needs manual reading/parsing of reference documents |
| No capability to evolve based on style patterns |

Without access to more synthetic resumes and open-source models, your current pipeline plateaus at 60–66% match, missing the fine-tuned richness and regulatory depth shown in manual resumes.  
 Unlocking more annotated samples + embedding model access is non-negotiable to cross the 90%+ accuracy threshold and scale across roles.

**4.Why More Samples Are Critically Important**

Even with a 97% format match, the manual content match remains only ~68%. Here's why more sample resumes (synthetic or real) are essential:

A. LLM Learns via Examples (Few-Shot Learning)

* The LLM imitates whatever structure/tone it sees in the EY sample resume.
* But it doesn’t know *what details matter per role* unless it sees many examples across domains: IFRS 9, TRIM, climate risk, IRB, NBFC, market risk, etc.
* One or two samples limit personalization. More resumes = richer representation of how actual humans express experience.

B. Manual Resume ≠ Just Format

* The manual resume includes:
  + Contextual phrasing (e.g., “ECL overlays applied to NBFCs and PSBs in stress scenarios”)
  + Domain insights (e.g., “Z-spread instruments validated under stressed VaR”)
  + Impact metrics, specific tools, and client outcomes
* LLM can only guess or generalize these unless you feed such styles through more samples.

| **5.Need** | **Why LLM is Required (Not Python-Only)** |
| --- | --- |
| **Sentence Rewriting** | LLM can **rephrase, merge, and summarize** points from resume into impactful, JD-aligned bullet points. Python cannot. |
| **JD-Specific Emphasis** | LLM adapts emphasis: if JD talks more about **SR 11-7**, LLM mirrors that. Python-only approaches lack such contextual weighting. |
| **Format & Tone Replication** | LLM mimics the **style and professional tone** from the EY sample format using prompt-based guidance. |
| **Domain Awareness** | LLM has learned **financial risk concepts** (e.g., PD, LGD, ECL, TRIM) and can arrange them meaningfully, unlike regex/text matchers. |
| **RAG-Based Personalization** | LLM + chunk retrieval enables **resume-JD matching** far better than fixed rules or keyword lists. |

**LLM Learns by Examples**

* You’ve only given 1 sample (EY\_ClientFacing\_Resume.txt)
* This limits the LLM to mimicking one tone/format
* In manual resumes, professionals make contextual choices — e.g., phrasing impact for NBFCs, overlays on PSBs, IFRS9 stress insights

**B. No Persona Differentiation**

* Manual resume reflects Kartik’s unique domain (e.g., IFRS9 for CRE portfolios)
* LLM lacks enough samples to learn how **GCB 6 vs GCB 4** phrasing differs
* Adding 9–12 resumes per GCB level (as per your chart) **enables tone differentiation**

**C. Missing Regulatory Domain Coverage**

* Without more JDs and resumes across **TRIM, IRB, CCAR**, etc., the model doesn’t adapt to subtle shifts in required phrasing
* For example: TRIM validation writeups look structurally different from IFRS9 or FRTB resumes

**D. Keyword Overlap ≠ Contextual Match**

* TF-IDF shows surface-level overlap
* But deeper context like **project outcomes, regulatory insight, modeling frameworks (Challenger, Benchmark, OPA)** are missing in LLM output without guided enrichment

More annotated examples allow the LLM to **learn how** to phrase these organically

**6.Steps That Require More Data + Exact Volume + Why It Matters**

| **Step** | **Data Needed** | **Exact Quantity Needed** | **Why It Improves 100% Customization & Match** |
| --- | --- | --- | --- |
| **1. Style Matching per GCB level** | EY resumes labeled with GCB (3, 4, 5, 6) | **3 per GCB level → 12 total** | Ensures tone, structure, seniority-based phrasing are replicated appropriately per role. Prevents over/under-framing of experience. |
| **2. JD → Resume Skill Mapping (Banking/Risk)** | Labeled JD-to-resume-mapped examples | **10-15 JDs + matched resumes** | Enables LLM to understand how experience is selectively surfaced to match key JD terms like “IRB”, “TRIM”, “SR 11-7”, “VaR”. |
| **3. Persona Differentiation** (Kartik ≠ Girish) | Manually customized resumes per person | **3 historical resumes per person** (for 3 personas = **9 total**) | Prevents repeated content. Learns unique tone and content structure per person (Girish = validation expert, Kartik = quant risk modeller, Anand = hybrid modeller). |
| **4. Domain Adaptation (Regulatory Risk / IFRS9)** | Different JD formats in same domain | **5+ JD variations in same domain (e.g. IFRS9, ECL, AIRB, CCAR)** | Expands the LLM's ability to align content with subtle variations in regulatory focus. Helps tailor bullet points precisely. |
| **5. Format Style Matching (Sample EY Profiles)** | EY-format resumes across teams | **5-7 distinct EY samples from FSRM, Credit Risk, Modelling, AI/ML** | Allows precise structural replication of team-specific expectations (e.g. modeling-heavy vs validation-heavy formats). |
| **6. Similarity Threshold Optimization** | Multiple versions of same resume (initial + final EY conversion) | **15–20 total resume-to-EY conversions** | Enables cosine similarity model to correctly evaluate what counts as “manual-level quality”. Trains matching thresholds. |
| **7. Keyword Coverage Validation** | List of “must-have” keywords with placement examples | **Just 1 labeled list + placement examples** | Ensures all resumes include EY-critical keywords. Helps enforce coverage logic (e.g., PD/LGD in Model Validation section). |

* **📊 Total Data Needed (Unique Files)**

| **Category** | **# Files Required** |
| --- | --- |
| JD–Resume pairs (diverse) | 15 |
| Manual EY-format resumes | 20 |
| Persona-specific resume versions | 9 |
| EY-format team samples | 5–7 |
| JD variations (within risk domain) | 5–7 |
| GCB-level labeled resumes | 12 |
| **Total (deduplicated)** | **≈ 40–45 unique files** |

**7.Why FAISS Is Not Working in My Code**

**✅ What's Causing the Errors?**

1. **SSL Certificate Verification Failed (SSLCertVerificationError)**
   * My Anaconda Python environment **cannot validate SSL certificates** from HuggingFace or regulatory websites.
   * Even after installing certifi, **the current Jupyter kernel doesn't pick it up**.
   * Likely cause: **EY’s corporate firewall or proxy** is blocking external HTTPS requests (very common).
2. **Model Download Fails**
   * FAISS RAG relies on sentence-transformers/all-MiniLM-L6-v2 from Hugging Face.
   * But the model:
     + **Isn’t cached locally**
     + **Can’t be downloaded** due to corporate firewalls.
3. **Skipped URLs**
   * URLs like analystprep.com, ssrn.com, and listendata.com **fail SSL verification** and are skipped during vector store creation.
   * FAISS still builds but with **limited domain knowledge** (only BIS, EBA, BoE succeed).

4.**What This Means**

* The FAISS vectorstore **does get created**, but it's **not enriched enough** for strong regulatory alignment.
* The model's summary and experience **match scores drop**, because **domain-specific context is missing**.

**8.Access Denied to Open-Source Models (LLM & Embeddings)**

* **The model mistral is locally hosted**, and no sentence-transformers or GPT model is used. This severely limits:
  + *semantic understanding*
  + *contextual paraphrasing*
  + *linguistic richness matching the sample resume style*
* **Embedding-based enrichment is not used**: Without Sentence Transformers like all-MiniLM-L6-v2, the system cannot semantically compare or extract context from similar resumes, limiting accuracy of customization.
* **No similarity filtering using advanced embeddings**: Since sentence-transformers couldn’t be used due to EY firewall or no internet, you're forced to rely on **TF-IDF** which is **surface-level** — it only captures term frequency, not semantic similarity.
* **No access to pre-trained resume fine-tuned LLMs** (like GPT-4, Claude, or open-source resume generators from Hugging Face), which would have better zero-shot formatting capabilities.

**9.Why Access Denied to Open-Source Models is the Second Core Blocker**

| Block | What’s Blocked | Impact |
| --- | --- | --- |

|  |  |  |
| --- | --- | --- |
| HuggingFace Models | Access denied to sentence-transformers, all-mpnet-base-v2, GPT-4, Mixtral, etc. | You are relying on basic TF-IDF + local Mistral without deep semantic understanding. |

|  |  |  |
| --- | --- | --- |
| Finetuning or Embedding Models | No option to pull semantic embedding models | Cannot build custom similarity scoring or embedding-based chunking. |

|  |  |  |
| --- | --- | --- |
| Scoring Models for Match Evaluation | Forced to use TF-IDF cosine similarity | Cannot use better semantic scoring like SBERT, OpenAI Embeddings, etc. |

|  |  |  |
| --- | --- | --- |
| LLM Prompting Power | LM Studio locally hosted model lacks few-shot, zero-shot power | LLM does not adapt dynamically to different tones or JD structures like OpenAI GPT-4 would. |

**10.Steps Used to Create EY\_ClientFacing\_Resume.txt from Anand, Kartik, and Girish Manuals**

**Skills Section Standardization**

* From each manual resume, a common set of **core skills** was identified:
  + IFRS9, SAS/SQL/Python, Model Validation, Credit Scorecards, Machine Learning, IRB
* Used checkmarks (✓) to denote proficiencies in each skill.

**3. Summary Section Generalization**

* Abstracted each individual's summary into a **generic format** that reflects:
  + Their **years of experience**
  + Their **specialization in model validation, regulatory frameworks (IFRS9, Basel), and risk management**
* Removed individual names or client-specific details to ensure **template usability**.

**4. Model Validation Section Aggregation**

* Combined **Model Validation** content from all three resumes.
* Structured the section using:
  + **Generalized phrasing** (e.g., “Conducted model validation”, “Performed benchmarking”)
  + Integrated **key terms** from all profiles: PD, LGD, EAD, ECL, AMR, OPA, SR 11-7, TRIM, RNIV
* Ensured bullets reflect:
  + Python/SQL automation
  + Documentation alignment
  + Governance standards
  + Model testing types: back-testing, overlays, benchmarking, performance diagnostics

**5. Other Experience Section Curation**

* Picked **common themes** from the non-validation experience:
  + SQL automation
  + Use of GCP/BigQuery (Girish)
  + RNIV & Z-spread models (Kartik)
  + Thought leadership articles (Kartik)

Included analytics and tool usage (Excel EUDAs, cubic spline, WoE/IV, EDA, etc.)

**💡 Why This Template Needs More Samples**

* The current version was built from **only 3 resumes**.
* Limitations:
  + Less diversity in phrasing and experience
  + Missing exposure to alternate formats, edge cases, or broader domains (e.g., Market Risk, Op Risk)
  + LLMs overfit on these specific phrasings, leading to **reduced personalization** and **limited adaptability**

✅ **With more samples**, we can:

* Create a **richer and more flexible prompt context**
* **Boost generalizability** across different JD-resume pairs
* Improve **match % with manual resumes** across diverse profiles

**Anand:**

Customized Version

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experience | | Model Validation across IFRS 9, MRM, Python-based validations, model monitoring for PSBs, etc. | | Expanded version of same with additional keywords like PD, LGD, ECL, RNIV, etc. added as per EY style |
| IFRS9 Models | Redeveloped and validated IFRS 9 models for Middle East banks | | Same content mapped and enriched under IFRS 9 PD/LGD/ECL model validation | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Macroeconomic | | Performed macroeconomic modelling for Middle East banks | | Executed macroeconomic forecasting and XGBoost/RF based ML models for defaults |
| Scorecards | | Developed scorecards for Indian and UAE banks | | Included scorecard development for PIL, USRB, ML-based models |
| **Regulatory Terms** | Mentions RBI and SS 11/23 explicitly | | Adds TRIM, SR 11-7, RNIV, VaR, Z-spread, OPA, AMR – to align with regulatory trends | |

|  |  |  |
| --- | --- | --- |
| ✅ Match with Girish manual resume | 68.12% | The generated resume partially captures personalization and specific domain details as in the manually customized version, but ~32% content is still generic or missed. |
| Summary | Experience in model validation and IFRS 9 (PD, LGD, EAD, overlays, challenger models, SR 11-7, TRIM) | 📝 Manual: 5+ years of experience, stronger focus on development + validation. ✅ Final: 3+ years of experience, focuses on validation, overlays, SR 11-7, TRIM, benchmarking. 📝 Manual: mentions regulatory capital frameworks and portfolio analytics. ✅ Final omits those explicitly but captures them in experience. |
| Model Validation | ✔ Both mention validation of IFRS 9 models, PD/LGD/EAD/ECL, use of Python, overlays, AUC/KS, documentation, macroeconomic overlays | 📝 Manual: Adds audit readiness, overlays documented methodology in detail, and includes impact assessment and assumption testing. ✅ Final: Focuses more on collaborative aspects (risk, audit, finance), and adds “model enhancement” and “robustness through challenger models.” |
| RNIV Validation | ✔ Both mention RNIV validation, Z-spread, VaR | 📝 Manual: Specifies assumptions testing, documentation, back-testing, MVR, and capital requirement calculations using Excel EUDAs. ✅ Final: Concise but accurate mention of validation including Z-spread. |
| Retail Scorecards | ✅ Both mention work on scorecards (PIL, USRB), validation lifecycle | 📝 Manual: No specific lifecycle breakdown. ✅ Final: Adds “full lifecycle from AMR to revalidation.” |
| Macroeconomic Forecasting | ✅ Mentioned in both | 📝 Manual: Macroeconomic risk factors described only in overlays for PD/LGD. ✅ Final: Adds forecasting models (GARCH) and defaults prediction using ML (XGBoost, RF) – a broader coverage. |

**🧠 Questions Testing Technical Depth**

1. *What exactly does the model do that a simple keyword-based script can’t?*
2. *Why are you using an LLM for such a basic task like formatting resumes?*
3. *Isn’t TF-IDF enough? Why bring in embeddings or RAG at all?*
4. *Can’t this just be done with a template and find-replace logic?*
5. *Why do you need FAISS and embeddings for something as straightforward as resume matching?*
6. *How does cosine similarity ensure relevance? Explain how it works in this context.*
7. *Why LM Studio? Why not OpenAI directly or another API?*
8. *How are you validating that the generated resume is actually better?*

**🤨 Questions That Undermine Your Judgment**

1. *Do you really think EY leadership will care about 5–10% match differences?*
2. *Isn’t this over-engineered for a task that's not client-facing?*
3. *Did you check if anyone else already solved this before reinventing the wheel?*
4. *How do you know this is aligned with EY brand and tone?*
5. *What’s the business impact of a 70% vs. 80% resume match? Can you quantify it?*
6. *How much time have you spent on this compared to other higher-priority tasks?*

**🕵️‍♂️ Gaslight-Framed ‘Innocent’ Queries**

1. *Are you sure you understood the MoM correctly? I don’t recall assigning both steps to you.*
2. *Did you actually clarify with Sapta or Nitin if LLM use is permitted here?*
3. *You’re saying access is blocked — did you really try every method or just gave up?*
4. *You’re blaming the lack of samples — or is it that your model isn’t good enough?*
5. *Were the manual versions even that different, or are you just trying to justify the output?*
6. *Do you feel you’re the best person to own this, given the dependencies?*

**💢 Aggressive, Challenging Questions**

1. *Why wasn’t this delivered earlier if it’s so automated?*
2. *Why did I have to hear about this through someone else?*
3. *Why should I trust the LLM-generated resume over a manually curated one?*
4. *If your automation is so good, why is the match only 65%?*
5. *Are you saying your model is smarter than the senior consultants writing manual resumes?*
6. *Is this output client-safe? Who validated this content?*

**🪤 Trap Questions to Corner You**

1. *So you’re saying access is blocked — why didn’t you escalate earlier?*
2. *You said more samples are needed — what stopped you from collecting them?*
3. *What exactly did you do that a fresher with GPT access couldn’t have done?*
4. *How did you test for hallucinations or compliance risks in LLM output?*
5. *Who approved this entire approach — or was it your individual decision?*

**🧾 Follow-ups to Force You to Doubt Yourself**

1. *Did you consider how risky this could be if the generated output went to a client?*
2. *Why didn’t you document your assumptions better?*
3. *You’ve written code — but did you peer-review or test its robustness?*
4. *Have you benchmarked this against any external standard or resume automation tool?*
5. *What makes your generated version better than the manual one by the SME?*
6. *Is the current process broken? Why are we even fixing it?*
7. *How do you measure "success" here? Who agreed to that metric?*
8. *Are you sure this isn’t more about overcomplicating things to stand out?*
9. *Are you saying the current team can’t do what you’re doing manually?*

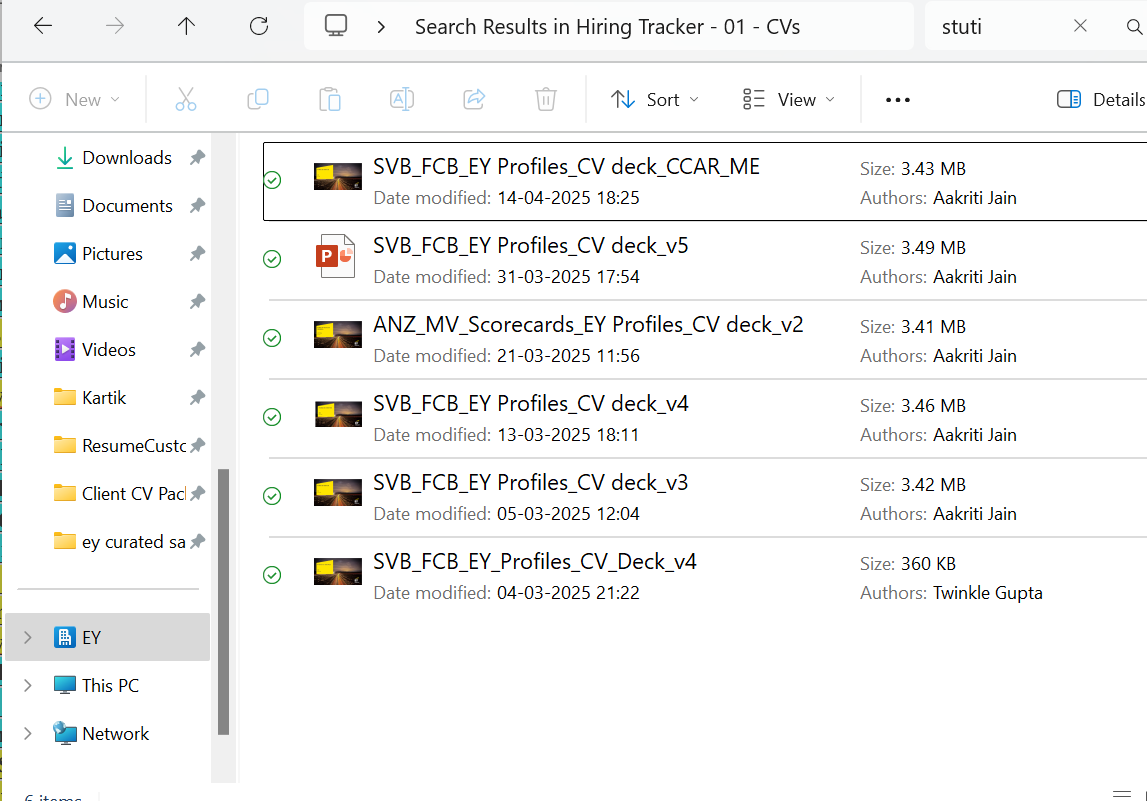
**🧠 How to Handle This Strategically**

Be ready with:

* Clear metrics (e.g., format match %, manual match %, time saved per resume)
* Your justification for using LLMs over rule-based scripts
* Highlight blocked access (open-source model unavailability, SSL firewall issues)
* Show side-by-side comparisons
* Stress repeatability, scale, and time savings
* Note that manual curation was used to build the reference sample, not ignored
* Always loop back to *"client-facing output consistency and EY brand alignment"*

Let me know if you want diplomatic responses or manipulation strategies for these.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **Name** |  |  |  |  |  |  |  |  | **Client Mapping** |  |  |  |  |  | **Raw Resumes** | **EY Curated Resume** |
| **1** | **Aarsh Shah** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **2** | **Sai Kiran** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **3** | **Reema Pandey** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **4** | **Nidhika Tomar** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **5** | **Arjun KM** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **No** | **Yes – HSBC , AIB** |
| **6** | **Rajvi Doshi** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **7** | **Kundan Kumar** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **8** | **Arnab Roy** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **9** | **Narendra Sahu** |  |  |  |  |  |  |  |  | **Lumen** |  |  |  |  |  | **Yes** | **Yes** |
| **10** | **Munna Chaudhary** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **Yes** | **No** |
| **11** | **Anand Kumar** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **Yes** | **Yes** |
| **12** | **Girish Khandelwal** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **Yes** | **Yes** |
| **13** | **Kartik Mohan** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **Yes** | **Yes** |
| **14** | **Shriwari Mhatre** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **No** | **Yes** |
| **15** | **Swar Sudha** |  |  |  |  |  |  |  |  | **HSBC** |  |  |  |  |  | **No** | **Yes** |
| **16** | **Rubika Rana** |  |  |  |  |  |  |  |  | **ANZ** |  |  |  |  |  | **No** | **Yes** |
| **17** | **Shashwata Mondal** |  |  |  |  |  |  |  |  | **ANZ** |  |  |  |  |  | **Yes** | **Yes** |
| **18** | **Kshitij Sahdev** |  |  |  |  |  |  |  |  | **ANZ** |  |  |  |  |  | **Yes** | **Yes** |
| **19** | **Vedanti Khokher** |  |  |  |  |  |  |  |  | **ANZ** |  |  |  |  |  | **Yes** | **Yes** |
| **20** | **Stuti Mehrotra** |  |  |  |  |  |  |  |  | **ANZ** |  |  |  |  |  | **Yes** | **Yes** |
| **21** | **Shambhu Dayal Singh** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **No** | **Yes** |
| **22** | **Shirwari Shedge** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **No** | **Yes** |
| **23** | **Vaibhav Yadav** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **24** | **Yash Rai** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **25** | **Susmita Misra** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **26** | **G N Sindhur** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **27** | **Abhimanyu Singh** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **28** | **Narendra Sahu** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **29** | **Debsubhra Ghosh** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **30** | **Anuska Das** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **31** | **Stuti Mehrotra** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **32** | **Praveen R** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **33** | **Rahul Kushwaha** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **34** | **Isha Porwal** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **35** | **Sahil Patil** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **36** | **Priyajit Bishayee** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **37** | **Subhadip Acharyya** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **No** | **Yes** |
| **38** | **Sulagna Paul** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **39** | **Sautrik Ganguly** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **Yes** | **Yes** |
| **40** | **Bhawya Mehra** |  |  |  |  |  |  |  |  | **SVB** |  |  |  |  |  | **No** | **Yes** |

****

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

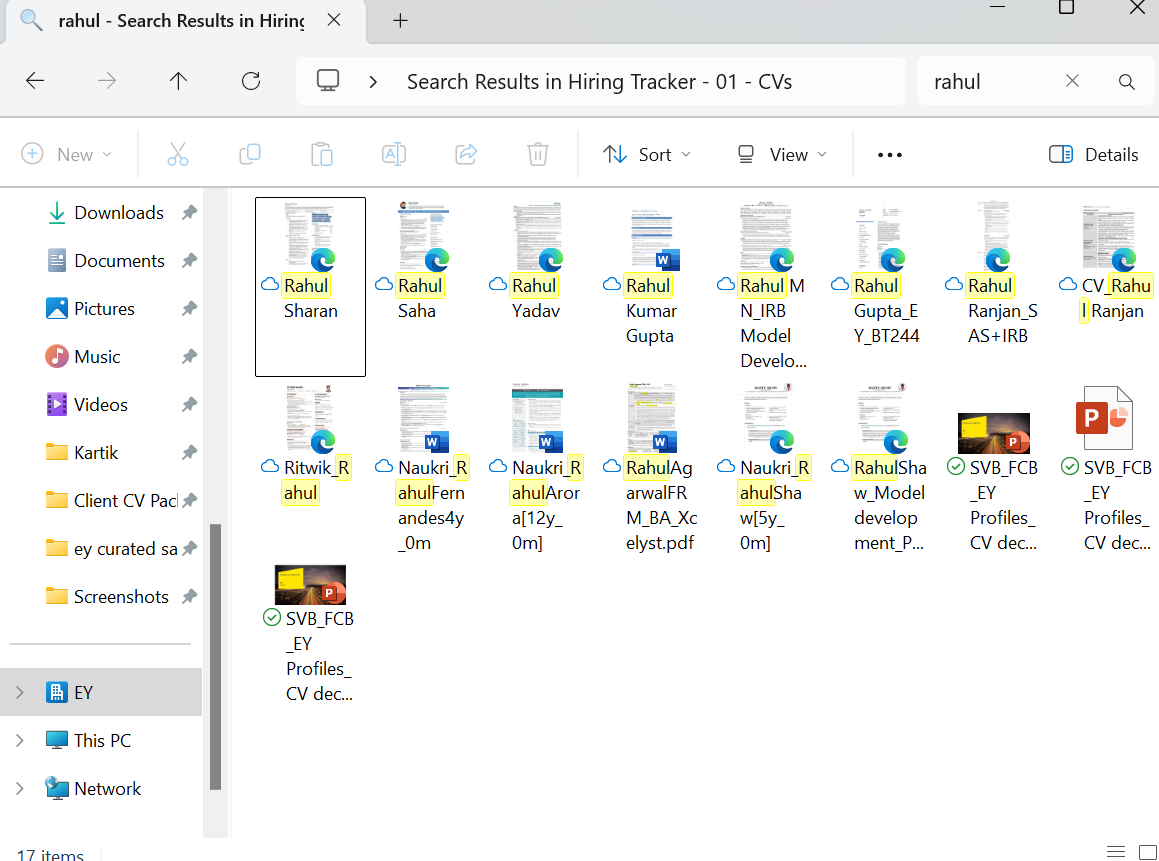
AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

****

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**List of Profiles in hsbcifrs9cvs**

| **S.No** | **Name** | **GCB Level** | **Key Highlights** |
| --- | --- | --- | --- |
| 1 | **Arjun KM** | GCB 4 | Basel Models, PCA Scorecards, AFT Model, PySpark, ML 101, Fair Lending |
| 2 | **Saumya Bhardwaj** | GCB 5 | IFRS9 SME PD/LGD, XGBoost, SAS, Python, SR11/7, CLI/CLD |
| 3 | **Ish Manchanda** | GCB 5 | IFRS9 & IRB validation, ECL error, TTC vs PIT, XGBoost |
| 4 | **Paarth Sharma** | GCB 6 | IRB & IFRS9, Vasicek, CRR, RTS, EBA/ECB audit, Uplift Models |
| 5 | **Shriwari Mhatre** | GCB 6 | Credit/Fraud, OOT/OOS, B-score, MLOps, Azure, Demand Forecast |

Hi I require JDs for anzey,anzscorecards,HSBCIMR,HSBCSWAR(this has only one candidate but on basis of which JD has her profile been customised?) ,SVBFCB-1,2,3,45.

Without JD,Our core objective of customising any resume to any JD is hindered . Since backtracing the customisation based on the manually customised resumes is redundant.

Also several overlapping of same candidates under different manual customised series- SVBFCB and HSBC.May I know what is the reason for the same and on what basis these overlap happen? If its based on JD requirements ,then once again please share the JDs as emphasised initially.

Hi,

To progress meaningfully on the core objective of mapping any resume to any JD, I require the following JDs at the earliest:

* ANZEY
* ANZ Scorecards
* HSBC IMR
* HSBC SWAR (this has only one candidate could you please confirm the specific JD used to customize her profile?)
* SVB FCB (Versions 1, 2, 3, and 4.5)

At this point, the absence of these JDs is significantly limiting the ability to scale or validate the customization pipeline. Reverse-engineering from manually customized resumes is increasingly redundant and unreliable, especially without visibility into the originating JD logic.

Moreover, I’ve observed overlapping candidate names across SVB FCB and HSBC manual resume series. Could you please clarify the reason for this duplication? If this was done based on differing JD alignments, then sharing those respective JDs becomes even more important to retain contextual accuracy.

**Is it possible to customize raw resumes *without* JDs and still achieve 100% match with manually customized resumes?**

**No, not fully — and definitely not 100%. Here’s why:**

**Why JD Is Essential**

1. **Manual customization is *driven by JD***:
   * Manual resumes are *tailored* to what a job expects: keywords, skills, responsibilities.
   * Without a JD, there’s **no context** to guide what to emphasize, retain, reword, or omit.
2. **LLMs and humans both need context**:
   * Even an LLM like GPT-4 or Mistral needs **goal signals** (JD) to produce output that *aligns with expectations*.
   * Without a JD, the model defaults to generic formatting, which will never match a manually *targeted* version.

**Why 100% Match Is Unrealistic Without JD**

| **Limitation** | **Explanation** |
| --- | --- |
| No target keywords | Manual resumes pick JD keywords like "PD/LGD/ECL", "VaR", "overlays", etc. |
| No role relevance | Manual customizations trim irrelevant experience and elevate what's needed. |
| No structural match | Manual resumes follow structure as per client expectations, which come from JDs. |
| Model guessing | LLM will "guess" what’s important, which will never be the same as human judgement guided by a JD. |

**🔑 What You *Can* Do Without JD**

If you're truly JD-less:

* Use a **synthetic JD or generic reference JD** (like an EY format one for a role)
* Train or fine-tune on past resume-JD pairs (synthetic data) to *approximate* match
* Aim for **70–80%** match, not 100%

**✅ Honest Conclusion:**

**You cannot achieve 100% match with manual customization without JD context because the very essence of manual customization is *JD alignment*.**

To achieve close match:

* Use a good JD (even synthetic)
* Use sample formats
* Leverage prompt engineering and LLM
* Validate using cosine similarity + keyword patching

**Why JD Access Is Absolutely Essential for Customization**

1. Every bullet, phrasing choice, and section emphasis in the manual version is carefully aligned to a job description. Without the JD, we’re working blind.
2. Trying to guess the original intent behind a manual resume is redundant and risks building incorrect patterns.
3. Without a JD, the LLM produces generic, non-targeted outputs that structurally look fine but fail to align with what the client or project expects.
4. Manual resumes selectively surface terms like “PD”, “LGD”, “ECL”, “VaR”, “TRIM”, “overlays” only because the JD called for them. Without that prompt, such precision cannot be replicated.
5. Manual customization trims away irrelevant experience and highlights what matters for that specific role. Without a JD, we don’t know what to elevate or exclude.
6. The order of sections, phrasing style, and tone in the manual version typically mirror expectations from the JD. That cannot be recreated in a vacuum.
7. Without the JD, the LLM "guesses" what’s important. That guesswork will never match the accuracy of a human who had clear JD context — making true replication impossible.
8. For this to work at scale, we must know what the manual version was responding to. The JD is that anchor.

**Immediate JD Access Required to Unblock Progress**

To move forward meaningfully, I urgently require the following job descriptions:

* ANZEY
* ANZ Scorecards
* HSBC IMR
* HSBC SWAR *(This has only one candidate. Could you please confirm which specific JD was used to customize her resume?)*
* SVB FCB (Versions 1, 2, 3, and 4.5)

At this point, **the lack of these JDs is significantly hindering both validation and scale** of the customization pipeline.

Hi ⁠Everyone I request you to address the following blocker at the earliest so that our target can be achieved.

Why JD Access Is Absolutely Essential for Customization

1. Every bullet, phrasing choice, and section emphasis in the manual version is carefully aligned to a job description. Without the JD**, we’re working blind.**
2. Trying to guess the original intent behind a manual resume is redundant and risks building incorrect patterns.
3. Without a JD, the LLM produces generic, non-targeted outputs that structurally look fine but fail to align with what the client or project expects.
4. Manual resumes selectively surface terms like “PD”, “LGD”, “ECL”, “VaR”, “TRIM”, “overlays” only because the JD called for them. Without that prompt, such precision cannot be replicated.
5. Manual customization trims away irrelevant experience and highlights what matters for that specific role. Without a JD, we don’t know what to elevate or exclude.
6. The order of sections, phrasing style, and tone in the manual version typically mirror expectations from the JD. That cannot be recreated in a vacuum.
7. Without the JD, the LLM "guesses" what’s important. That guesswork will never match the accuracy of a human who had clear JD context — making true replication impossible.
8. For this to work at scale, we must know what the manual version was responding to. The JD is that anchor.

Immediate JD Access Required to Unblock Progress

To move forward meaningfully, I urgently require the following job descriptions:

* ANZEY
* ANZ Scorecards
* HSBC IMR
* HSBC SWAR (This has only one candidate. Could you please confirm which specific JD was used to customize her resume?)
* SVB FCB (Versions 1, 2, 3, and 4.5)

At this point, the lack of these JDs is significantly hindering both validation and scale of the customization pipeline.

Hi Nimish,

First, thank you sincerely for sharing the ~40 raw resumes along with their manually customized versions — that’s a significant input, and I truly appreciate it. It will help improve model performance across a wider set of personas and banks, and the progress so far is encouraging.

That said, we’re now at an inflection point. These resumes span distinct client contexts - ANZ, SVB, Lumen apart from HSBC, yet without their originating JDs, the LLM is forced to do “guesswork”. Even with full JD access (e.g., for Kartik, Girish, Anand), match scores peak around 65–78%. Without JD context, that score will likely drop sharply due to incorrect keyword prioritization, tone misalignment, and loss of role-specific structuring. One-liner summaries like “CCAR” or “IFRS9” aren’t enough the model can’t replicate manual logic from titles alone, and reverse engineering becomes unreliable.

To proceed meaningfully, I urgently need the JDs for ANZEY, ANZ Scorecards, HSBC IMR, HSBC SWAR (which has only one candidate - kindly confirm the JD used), and SVB FCB (Versions 1–5). While I’m happy to go ahead using the HSBC IFRS9 JD for broader coverage, it would be technically inaccurate for non-HSBC resumes and will likely degrade similarity scores. I’m also happy to get on a call with Aakriti and co-create JDs using EY-Q — but her domain expertise will be absolutely essential to ensure accuracy.

Warm regards,  
**Akshara**

**Subject:** JD Access – Critical for Ensuring Customization Accuracy and Preventing Pipeline Deviation

Hi Nimish,

First off, thank you sincerely for sharing the **~40 raw resumes along with their manually customized counterparts**. This level of access is incredibly helpful and appreciated — it clearly reflects the thought and effort invested in building this repository, and it has significantly accelerated our progress on model tuning.

That said, we are now at a point where **further improvement and scalability depend entirely on JD access** — and I wanted to outline why, both from a modeling and quality assurance standpoint.

**Why JD Access Is Now Critical**

1. **Manual resumes were never created in isolation — they are JD-driven.**  
   Every line, tone shift, and keyword is shaped in response to a job’s requirements. Without the JD, we lose the core logic behind these editorial decisions.
2. **The ~40 resumes provided span ANZ, SVB, Lumen — all with distinct client needs.**  
   While I can technically apply the HSBC IFRS9 JD to these, it will not reflect the actual intent behind the customizations. This risks **forcing the model into false assumptions**, resulting in poor alignment and degrading score accuracy.
3. **Without the correct JD, I cannot ensure match fidelity.**  
   The LLM may hallucinate or overfit to the wrong structure — for instance, applying IFRS9 language to CCAR or scorecard roles — producing technically sound but strategically off-target resumes.
4. **Similarity scores plateau even when JDs *are* provided.**
   * Kartik: **78.18%**
   * Girish: **69%**
   * Anand: **65.55%**  
     These numbers were achieved using full JD access. Without that baseline, scores can drop below 50%, undermining confidence in model performance.
5. **LLMs, like humans, require JD signals to produce role-specific alignment.**  
   In the absence of JDs, the model defaults to generic phrasing. This means:
   * Loss of critical terms like PD, LGD, CCAR, VaR, overlays
   * Incorrect tone or prioritization
   * Structure that doesn’t match client-specific expectations
6. **One-liner intent summaries (e.g., “macro scenarios”, “scorecards”) aren’t enough.**  
   These lack the technical granularity and contextual clues that drive the nuance of a high-fidelity customization. The model can’t infer JD tone, hierarchy, or emphasis from these labels alone.
7. **Reverse-engineering manual edits is now counterproductive.**  
   Attempting to guess what JD a manual resume was written for — without visibility into it — leads to inconsistent training signals and increases the model's error rate.

**What I Need to Proceed Confidently**

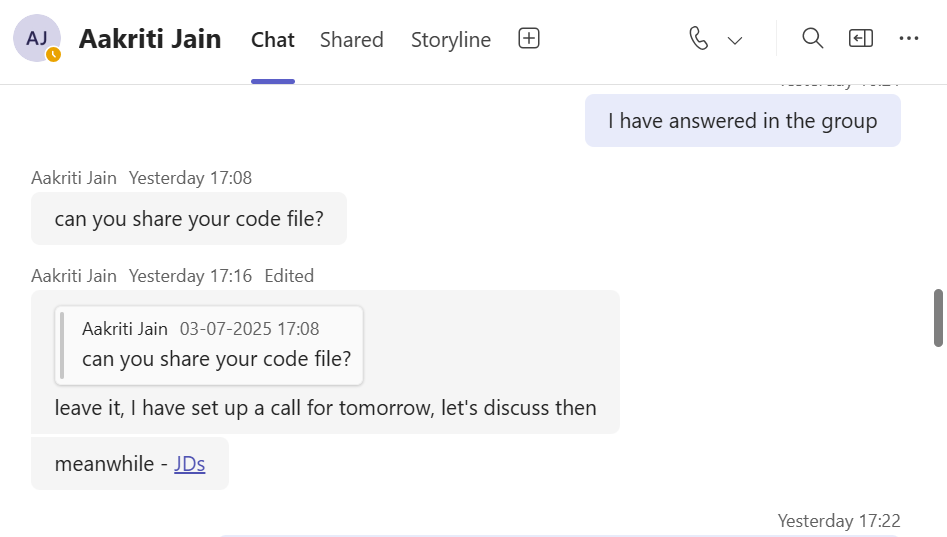
To ensure quality and move toward scalable automation, I kindly request the JDs (or request notes/near-final versions) used for:

* **ANZEY**
* **ANZ Scorecards**
* **HSBC IMR**
* **HSBC SWAR** *(Only one candidate exists — could you confirm the JD that drove her customization?)*
* **SVB FCB (Versions 1, 2, 3, and 4.5)**

I want to reiterate that I’ve always been open to collaborating and aligning with you. That’s precisely why I proactively requested a call not just once, but multiple times across the personal message,group and via email with the sole intent of ensuring we’re aligned on the JD requirements and execution plan.

I didn’t proceed with the 4 PM call as I sensed a strong resistance from you to my request, despite my sharing specific reasons around the need for JDs. I’d like to gently share that as someone newer to this setup. I’ll be honest while I’ve felt quite at ease collaborating with everyone else across the board, I did find your interactions to be intimidating me .Please try to have basic empathy and it’s a sincere request. At the same time, I fully acknowledge your expertise and contributions to this space, and I am only trying to ensure the best outcome for the task at hand.

If an escalation has been made, I’d appreciate some transparency around it, including who was looped in. That said, I consider it my responsibility to keep relevant stakeholders informed not as a means to sidestep alignment, but to ensure we’re moving with full visibility. I continue to remain fully committed to collaborating with you constructively and respectfully as we take this important workstream forward.



----- Abhimanyu Singh svbfcb3.txt -----

Abhimanyu Singh

Role: Senior

Summary:

Abhimanyu is a Senior in EY's Financial Services Risk Management practice.

He has 3+ years of experience in Banking and Financial Services.

Proficient in CCAR/DFAST model development and stress testing under CCAR/CECL norms.

Qualifications and Training:

Integrated M.Sc. in Geology, IIT Kharagpur, 2021

Technical Skills:

Basel II/III, CCAR, PPNR, IFRS9 frameworks

Credit risk model development & validation

Stress Testing, Capital Policy & Planning, Risk Analytics

Advanced Excel, SAS, SQL, Tableau, Jira

Key Relevant Experience:

Built PD model using hazard functions, behavioral & macro variables.

Designed PPNR forecasting models for CCAR stress testing.

Led macroeconomic monitoring and reporting initiatives.

Validated CECL loss models using discrete-time survival analysis.

Improved risk model compliance via data reconciliation and risk factor recalibration.

----- Anand Kumar hsbcifrs9v2.txt -----

2. Anand Kumar

M.Sc. in Statistics | GCB 5

Summary

Anand has 10+ years of experience in statistical analysis and credit risk modelling for global banks and Indian PSBs. He has deep expertise in both model validation and development, leveraging Python in credit risk and regulatory capital frameworks.

Model Validation

Validated IFRS9 model for Middle East banks using Python; subsequently passed external audits.

Conducted full validation (assumption testing, benchmarking) for IFRS9 retail models.

Authored validation reports aligned with MRM standards (SS 1/23).

Created Python-driven monitoring templates for Indian PSBs as per RBI.

Led model monitoring for Middle Eastern institutions – performance drift detection, visual reports.

Other Experience

Led scorecard development for PSU and Middle East banks using R, Excel, and SAS.

Redeveloped IFRS9 models for Middle East banks under SAMA guidelines.

Built machine learning models for delinquency prediction (SAS, R).

Developed scorecards per CBUAE norms for UAE and Indian banks.

----- Anand Kumar hsbcimr.txt -----

Anand Kumar

M.Sc. in Statistics

GCB 5

Summary

Anand has more than 10 years of experience in statistical analysis and credit risk modelling for global banks and Indian PSBs. He has deep expertise in both model validation and development, leveraging Python to deliver robust solutions across credit risk domain and regulatory capital frameworks.

Experience

Model Validation

Validated redeveloped IFRS 9 model for Middle East banks; automated analysis and documentation using Python

Validated by external auditors; strong IFRS 9 and retail credit risk knowledge

Used Python to replicate model logic, test assumptions, benchmark performance

Documented validation reports for European bank (SS 1/23 compliance)

Designed/implemented Model Validation & Monitoring framework for Indian PSBs per RBI guidelines

Conducted model monitoring in Middle Eastern institutions (drift detection, stability, visual reporting)

Other Experience

Led scorecard development for PSU and Middle East banks (R, SAS, Excel)

Redeveloped IFRS 9 models per SAMA guidelines

Macro-economic modelling using R and Excel

Built ML models to predict bad accounts (R, SAS)

Developed scorecards per CBUAE for Middle East and Indian banks

----- Anuska Das svbfcb4.txt -----

4. Anuska Das – Senior

Summary:

4 years in BFSI with experience in CCAR and CECL execution and UAT.

Qualifications:

MSc & BSc in Applied Economics, Presidency University (2020 & 2018)

Key Experience:

Executed CCAR and climate stress tests (SQL, Excel).

CECL model execution using Python (PD, LGD, EAD).

UAT of CCAR platforms.

Backtesting and monitoring for portfolio quality.

Migration projects using PySpark.

Technical Skills:

SAS, Python, SQL, Excel, VBA

CCAR, CECL execution

Climate stress testing, backtesting

----- Arjun KM anzey.txt -----

Senior Manager: Arjun KM

Summary

Arjun is a part of the Financial Services Risk Management practice of EY India with 10 years of experience. He specializes in credit risk modelling, treasury and marketing analytics, scorecard development, and has proficiency in machine learning, deep learning, statistical modelling, and programming.

Qualifications and training

Masters - Statistics and Machine learning, Indian Institute of Science, Bangalore

B Tech - Mechanical Engineering, School of Engineering, CUSAT

Key relevant experience

IFRS 9 ECL modelling (time series, segmentation, macroeconomic analysis)

Basel Modelling: PD, LGD, EAD models for Auto business RWA

Reviewed & validated PD quantification model for US bank

Monitored production models & portfolio exercises (Business, Governance, Audit)

Built PCA-based scorecard for MFI Centre

Developed Adaboost classifiers for delinquency prediction

Created MFI Risk scorecard for MFI Bureau customer evaluation

Built ML-based warning indicators for large US bank health monitoring

Built Fair Lending compliance framework (logistic/linear regression)

Developed Digital Index (score-based segmentation) for US retailers

Built Naïve Bayes classifiers for enrolment prediction

Proprietary Recommendation Engine for US retail clients

Built predictive analytics models for aircraft engine maintenance

Led PySpark modelling ecosystem development with Bokeh dashboards

Built Random Forest models for shopper prediction

Designed ML 101 certification; trained over 300 employees

Technical Skills

R, Python, SQL, C, Git, Excel

EDA, AI/ML modelling

Credit risk model validation

MS Office

----- Arjun KM hsbcifrs9cvs.txt -----

Arjun KM

Masters – Statistics and ML | GCB 4

Summary

Arjun is a part of the Financial Services Risk Management practice of EY India with 10 years of experience. He specializes in credit risk modelling, treasury and marketing analytics, scorecard development, and proficiency in machine learning, deep learning, statistical modelling, and programming.

Experience

Risk Modelling and Analytics:

Pioneered Basel Modelling initiatives: developed PD, LGD, and EAD models for Auto business RWA calculation

PCA-based scorecards for MFI Centre risk assessment

Adaboost classifiers for customer delinquency prediction

MFI Risk scorecard to evaluate customer risk

Survival & AFT models, CAP segmentation, heuristic optimization

Aircraft engine maintenance prediction using Time Series & Regression

Data Science and ML:

Led PySpark modelling ecosystem with Bokeh dashboard

Delivered ML 101 to 300+ employees in 4 years

Random Forest classifiers for multichannel shopper prediction

Proprietary recommendation engine for US retail clients

ML models for US bank health and Fair Lending compliance

Other:

Reviewed and replicated PD models

Involved in model monitoring, stakeholder management

Conducted research on novel analytics techniques

----- Debsubhra Ghosh svbfcb4.txt -----

3. Debsubhra Ghosh – Manager

Summary:

5+ years in BFSI with focus on PPNR, CCAR models, credit risk scorecards.

Qualifications:

M.Sc. Statistics, Ballygunge Science College (2017)

B.Sc. Statistics, RKM Residential College (2015)

Key Experience:

PPNR model development for 9-quarter forecasts (AUM, AUC).

Revenue forecasting for Investment Banking under macro stress.

Worked on CECL and CCAR with PD, LGD, EAD models.

Built scorecards using logistic regression, reject inference.

Created automation tools for WOE binning and data quality.

Technical Skills:

SAS, Python

Time Series Analysis

Credit Risk Scorecards

MS Office

----- Debsubhra Ghosh svbfcb5.txt -----

Debsubhra Ghosh

Role: Manager

Summary:

Debsubhra is a seasoned professional in EY’s Financial Services Risk Management practice.

He has 5+ years of experience in PPNR, CCAR models, credit risk scorecards.

Qualifications and Training:

M.Sc. in Statistics, Ballygunge Science College, 2017

B.Sc. in Statistics, Ramakrishna Mission Residential College, 2015

Key Relevant Experience:

Developed PPNR models to forecast 9-quarter balances under macroeconomic scenarios

Led revenue forecasting for Investment Banking division

Managed CCAR/CECL stress testing for retail banking

Developed scorecards using logistic regression

Built automation tools in SAS and PySpark

Technical Skills:

PPNR, CCAR, CECL, Scorecards

SAS, Python, MS Office

----- G. N. Sindhur svbfcb3.txt -----

G. N. Sindhur

Role: Senior

Summary:

Sindhur is a Senior in EY’s Financial Services Risk Management practice.

He has 5+ years of experience in the financial services sector.

Specializes in Model Development & Validation to enhance risk management processes.

Skilled in econometrics modeling, PD, EAD models under CCAR and CECL frameworks.

Qualifications and Training:

M.Sc. Economics, Indira Gandhi Institute of Development Research, 2020

PGDM, T.A. Pai Management Institute, 2017

B.E. (Hons) Mechanical Engineering, BITS Pilani Hyderabad, 2013

Technical Skills:

CCAR & CECL model development (PD, EAD)

Econometric Model Development, Quantitative Finance Analysis

Tools: R, EViews, Stata, Python, Advanced Excel

Key Relevant Experience:

Developed CCAR PPNR models for capital planning using stress scenarios.

Built and validated credit risk models (PD, EAD) across North American credit card portfolios.

Supported IFRS9 ECL system implementation for Indian banks.

Led data science projects in financial risk management.

Conducted model validation, back-testing, documentation, and compliance checks.

Automated data extraction for CECL monitoring.

----- Girish Khandelwal hsbcimr.txt -----

Girish Khandelwal

MBA in Banking & Management

GCB 5

Summary

Girish holds more than 10 years of experience in the risk management space for banking clients. He specializes in Credit Risk, Model Monitoring, Validation & Development, Basel & IFRS9 Regulatory Reporting, Business Data Analytics, and Collection Strategy.

Experience

Model Validation

Independent validator for credit risk models; ensured governance/regulatory compliance

Python-based back testing, benchmarking, stress testing

Led scoring team (PIL, USRB) for monitoring/validation/development

Basel III, IFRS 9, CCAR, scorecard models (Acquisition, ECM, Collections)

Revalidation, AMR, OPA using FICO, ISPL, Experian scorecards (Python analytics)

PD/LGD/EAD validation for credit risk management

Managed full model lifecycle (1st line of defence); contributed to MRM guidance (SR 11-7)

Other Experience

SAS & SQL for reporting retail portfolio performance

Google BigQuery & GCP for data analysis

Automated shock simulation for IFRS9/IRRBB using SQL

----- Ish Manchanda hsbcifrs9cvs.txt -----

Ish Manchanda

MSc in Economics | GCB 5

Summary

Ish is a Manager in EY FSRM India with 4+ years of credit risk experience. He has worked in model development, validation, and audit.

Experience

Model Validation:

IFRS9 validation for UK bank (Retail UK & India)

Performed methodology review, PD/LGD backtesting, SAS code replication

ECL materiality and estimation error assessments

XGBoost marketing model validation for US bank (hyperparameter tuning, Gini/accuracy)

Other:

Qualitative/quantitative IFRS9 PD reviews (MENA mid-market banks)

Reviewed TTC PD, PIT macroeconomic modeling

IRB PD remediation for EU mortgage bank

Data augmentation and MoC using SAS

----- Isha Porwal svbfcb7.txt -----

Isha Porwal

Role: Senior

Summary:

Isha is a Senior in EY’s Financial Services Risk Management practice with around 4 years of BFSI experience.

Qualifications and Training:

PGDM (Banking & FS), NIBM Pune, 2021

BBA, IIPS DAVV Indore, 2019

Key Relevant Experience:

Delivered PPNR projections for FINMA Loss Potential & Risk Appetite

Forecasted revenue under macro stress, presented to CFO/COO

Executed 80+ PPNR models

Participated in ESAF Governance Forum

Reconciled LPA & Pillar 3 disclosures

Assessed MSME loans (up to ₹20 Cr) using Perfios, Karza, Highmark

Evaluated credit risk, legal, and technical parameters

Technical Skills:

PPNR Modelling & Execution

IFRS9, ICAAP

Python, R, SPSS, Stata, MS Office

----- Jessica Thacker svbfcb3.txt -----

Jessica Thacker

Role: Manager

Summary:

Jessica is a Manager in EY's Financial Services Risk Management practice.

She has 6+ years of experience in the Banking and Financial Services industry.

Proficient in CCAR model development and data analytics.

Qualifications and Training:

PhD Economics, South Asian University

MA Economics, South Asian University

BA Hons (Economics), University of Delhi

Technical Skills:

Statistics, Econometrics (Cross-Section, Panel Data, Non-Linear Regression, Time-Series)

Machine Learning (CART, Bagging, Boosted Trees)

Statistical Software: Stata, SAS, MS Office

Programming: Python, R

Key Relevant Experience:

Led PPNR CCAR econometric model development for revenue projections (SR15-18 compliant).

Conducted ongoing monitoring and back-testing of PPNR models.

Managed data analysis, variable selection, econometric strategy, and robustness checks.

Initiated machine learning-based revenue forecasting (Random Forest, XGBoost, CART).

Applied econometric and ML methods to predict firm performance, distress, and productivity.

----- Kartik Mohan hsbcimr.txt -----

Kartik Mohan

MA in Economics

GCB 6

Summary

Kartik has 5+ years of experience in statistical analysis and credit risk modelling for global banks. He specializes in model validation and development using Python with focus on credit/market risk models and regulatory frameworks.

Experience

Model Validation

IFRS 9 model remediation/validation (PD, LGD, EAD, ECL)

Reviewed development documentation; executed validation testing

Backtesting and benchmarking against realized performance

Stability checks, discriminatory power (AUC, KS), data quality

Applied overlays (macro risk factors) for CRE portfolio

Documented overlays, assumptions in Python

Validated RNIV market risk models (assumptions testing, benchmarking, reporting)

Other Experience

Backfilling model for inflation-linked interest rate curves (Python)

RNIV Z-spread models, EUDA Excel tools for capital requirement

PD model dev (logistic regression, Python) for US consumer portfolio

Authored article on MRM challenges (TRIM, FRTB, SR 11-7)

----- Ketaki Sardeshpande hsbcifrs9v2.txt -----

4. Ketaki Sardeshpande

MA Economics | GCB 6

Summary

Ketaki is part of EY FSRM with 3+ years of experience in credit risk model development and validation. She has worked across the UK and Ireland on development, validation, and audit teams.

Model Validation

Validated IFRS9 retail models for UK bank – methodology review, testing, sensitivity analysis (Python, SAS).

Independently replicated SAS portfolio monitoring, validated with Python.

Performed model testing (Gini, KS, PSI, correlation).

Built challenger EAD model for credit cards using CCF and linear regression in Python.

Validated IRB & IFRS9 PD/LGD models for UK challenger bank – calibration, scorecard methodology, accuracy.

Other Experience

Contributed to IRB LGD redevelopment for Irish bank using Python (RDS, EDA, SFA/MFA, calibration, MoC).

Conducted testing (homogeneity, stability, grade concentration), authored Model Dev Docs.

Used Python for default definition impact, MoC recalculations, JST/IVU-aligned PD model remediation.

----- M.V. Kailash Reddy hsbcifrs9v2.txt -----

1. M.V. Kailash Reddy

B Tech., Mechanical Engineering | GCB 5

Summary

Kailash is a part of the Financial Services Risk Management practice of EY India with 5+ years of experience. He has expertise in credit analytics, data analytics, data science and risk analytics/decision sciences spanning multiple industries like Banking, Insurance.

HSBC Experience

Recalibration/redevelopment of scorecard models for HSBC retail portfolios (Credit Card, Personal Loans – EMEA region) using SAS.

Collaborated with HBUK MMR team to support SS1/23 guideline implementation, perform PMA reviews, and generate RWA impact assessment reports.

Model Validation

IFRS9 LGD model validation for mortgages across Irish Bank geographies.

Designed monitoring metrics (PSI, transfer criteria, PD/Loss curves, LGD, ECL) using Python.

Conducted IFRS9 model reviews for Middle East banks including desktop review, reperformance, sensitivity testing, and benchmarking.

Other Experience

Developed IRB LGD model for a UK mortgage portfolio (data prep, EDA, calibration, testing, documentation).

Used Python for performance testing – homogeneity, HHI, PSI, Spearman, AUC, etc.

Built ELBE model for Irish bank’s IRB LGD using Python.

IRB EAD and PD remediation for a UK bank.

----- Manager Nidhika Tomar anzscorecards.txt -----

Manager: Nidhika Tomar

Summary:

Nidhika is a Manager in EY’s Financial Services Risk Management practice. She has more than 5 years of experience in the Banking and Financial Services industry. She is proficient in model development and implementation of credit risk models with exposure to regulations like Stress Testing, ICAAP, etc.

Qualifications and training:

Masters in Economics, Delhi School of Economics, 2019

Bachelors in Economics, Delhi University, 2017

Key relevant experience:

Worked on model development of Stress-testing portfolio evolution model (s166 secured book-shape model) which feeds into the capital (RWA/EL) and impairment (ECL) engines to predict the forecast compliant to the most recent usage of ICAAP across base/stress

Developed Capital suite (capital emulator- PD, EAD, LGD) implemented on portfolio evolution model

Engineered a comprehensive Impairment Emulator integrating Forward-Looking PD, EAD, LGD

Added sub-models like PPD and FSD to improve accuracy of impairment forecasts

Implemented stress-testing modelling using PyCharm

Developed BUK Home Finance Hybrid PD model

Recalibrated income models for unsecured loans and overdraft facilities

Guided 1.2 mn accounts’ affordability (£6bn exposure)

Investigated HPI changes’ impact on collateral value

Monitored RWA requirement for 1mn accounts (£150bn exposure)

Technical Skills:

PD/LGD Model Development, Stress Testing, ICAAP, RWA Computation

SAS, SQL, Excel, PyCharm, Python, GIT & MS Office

----- Munna Choudhary hsbcifrs9v2.txt -----

3. Munna Choudhary

MSc in Finance | GCB 5

Summary

Part of EY FSRM with 5+ years of experience in Credit Risk Modelling, IFRS 9, ECL estimation, financial forecasting, portfolio analytics, stress testing, and predictive modelling for business strategy.

Model Validation

Validated UK-based decision and behavioural models using Python – discriminatory power, trend analysis, RAG status.

Contributed to SS1/23 MRM efforts for a GSIB.

Built monitoring infra with PSI, CSI, Gini using Python.

Other Experience

Led IFRS9 PD/LGD/EAD model development for SME Real Estate and HNI Classic Car portfolios (15% improvement).

Designed Score Lumping algorithms for model efficiency.

Recalibrated Through-the-Cycle PD using Pluto Tasche approach.

Developed macro forecasting models to assess borrower behavior.

Supported ICAAP via stress testing, sensitivity analysis, PRA-aligned risk assessments.

----- Narendra Sahu svbfcb3.txt -----

Narendra Sahu

Role: Senior Manager

Summary:

Narendra is a Senior Manager in EY’s Financial Services Risk Management practice.

He has 10+ years of experience in the Banking and Financial Services industry.

Seasoned financial modeler with expertise in automating reporting and data controls, developing CCAR stress testing models, and implementing ECL calculations in compliance with IFRS9 standards.

Qualifications and Training:

IIM Executive Program: Fintech, Banking and Applied Risk Management, 2023

Madras School of Economics: MSc Financial Economics, 2017

Veer Surendra Sai University: Electrical Engineering, 2014

Technical Skills:

R, R-shiny, R-Markdown, Python, VBA, SQL, SAS, Git, Excel

Tableau, Power BI, AWS, Azure

Machine learning, AI

Statistical modelling, Data Analysis & Visualization, BCBS239

Time series, scenario analysis, IFRS9 assessment, PiT PD, PiT LGD, PiT EAD & CCF

Leadership, Client management, Agile, Scrum

Key Relevant Experience:

Developed and monitored CCAR stress testing models, participated in BAU reporting.

Built IRB models (PD, LGD, EAD, CCF) per regulatory guidelines using statistical techniques.

Helped a European bank with BCBS239 transition.

Built R Shiny-based ECL UI tools for Middle Eastern banks.

Created credit scoring models for a major Myanmar bank using Python.

Designed behavioral models with Random Forest.

Built exposure-based climate risk models using Bayesian networks and Monte Carlo simulations.

Led automation using R Markdown & Shiny, removing manual steps.

Delivered CCAR-compliant end-to-end credit model validations.

----- Narendra Sahu svbfcb4.txt -----

2. Narendra Sahu – Senior Manager

Summary:

10+ years in BFSI with expertise in CCAR stress testing, ECL (IFRS9), IRB models, and automation.

Qualifications:

Executive Program, IIM (2023)

M.Sc. Financial Economics, Madras School of Economics (2017)

B.E. Electrical, Veer Surendra Sai University (2014)

Key Experience:

Built CCAR stress testing models, quarterly BAU reporting.

IRB (PD, LGD, EAD, CCF) model calibration.

ECL app in R Shiny (Middle Eastern banks).

Credit scoring for a Myanmar bank (Python).

Climate risk modeling using Bayesian networks.

Model deployment, UAT, automation via R Markdown, R Shiny.

Delivered end-to-end model validations.

Technical Skills:

Python, R, R-Shiny, SQL, VBA, SAS, Git, Excel

Power BI, Tableau, AWS, Azure

Machine learning, scenario analysis, BCBS239

----- Narendra Sahu svbfcb5.txt -----

Narendra Sahu

Role: Senior Manager

Summary:

Narendra is a Senior Manager in EY’s Financial Services Risk Management practice. He has 10+ years of experience in the Banking and Financial Services industry.

Seasoned financial modeler with expertise in automating reporting and data controls, developing CCAR stress testing models, and implementing ECL calculations in compliance with IFRS9 standards.

Qualifications and Training:

IIM Executive Program: Fintech, Banking and Applied Risk Management, 2023

Madras School of Economics MSc: Financial Economics, 2017

Veer Surendra Sai University: Electrical Engineering, 2014

Key Relevant Experience:

Model Development and validation experience:

Development and monitoring of stress testing CCAR models.

Developed stress testing models using Python.

Developed IRB models (PD, LGD, EAD, CCF).

Worked with European bank on BCBS239 compliance.

Built ECL UI application using R Shiny.

Developed credit scoring models for Myanmar bank.

Built behavioral models with Random Forest.

Developed climate risk models with Bayesian networks and Monte Carlo simulation.

Created R Shiny-based UI tools.

Led automation initiatives in R Markdown and R Shiny.

Delivered full-cycle model validations and documentation.

Technical Skills:

R, R-shiny, R-Markdown, Python, VBA, SQL, SAS, Git, Excel

Tableau, Power BI, AWS, Azure

Machine learning, AI

Statistical modelling, Data Analysis & Visualization, BCBS239

Time series, scenario analysis, IFRS9, PiT PD, PiT LGD, PiT EAD & CCF

Leadership, Client management, Agile, scrum

----- Nidhika Tomar svbfcb5.txt -----

Nidhika Tomar

Role: Manager

Summary:

Nidhika is a Manager with ~5 years of BFSI experience. Specialized in PD/LGD model development, stress testing, RWA, and ICAAP.

Qualifications and Training:

Masters in Economics, Delhi School of Economics, 2019

BA Economics, Delhi University, 2017

Key Relevant Experience:

Built secured portfolio evolution models feeding into RWA/ECL

Developed capital emulator suite (PD, EAD, LGD)

Engineered impairment emulator with FL overlays (PPD, FSD)

Implemented modeling using PyCharm

Built hybrid PD model for BUK Home Finance

Assessed affordability on 1.2M accounts (~£6B exposure)

Measured HPI impact on £150B property value

Managed RWA tracking for internal/external stress tests

Technical Skills:

SAS, SQL, PyCharm, Python, GIT

Modeling: PD/LGD, Stress Testing, ICAAP, RWA

----- Paarth Sharma hsbcifrs9cvs.txt -----

Paarth Sharma

PGP in Data Science and Engineering | GCB 6

Summary

Paarth is with EY FSRM India, having 6 years of experience in IFRS9/Basel IRB model development, validation, and audit across multiple geographies.

Experience

Model Validation:

IFRS9 PD/LGD/Stress Test models – UK bank

IRB PD/LGD model validation (qualitative + quantitative)

Stress Test validation (uplift, default, DD shift by scenario)

IFRS9 PD validation (Middle East bank) using Python and SAS

Other:

IFRS9 PD development (Vasicek framework – Middle East bank)

Audit of IFRS9 PD/LGD/EAD models

IRB use-test audit for UK bank (CRR, RTS, EBA, ECB compliance)

IRB PMA control and substantive testing audit

----- Praveen R svbfcb5.txt -----

Praveen R

Role: Manager

Summary:

Praveen is a Manager in EY’s Financial Services Risk Management practice. He has 7+ years of experience in the Banking and Financial Services industry.

Experienced in stakeholder management, CCAR model validation, execution and analysis of CCAR, IFRS9, scorecards and rating models.

Qualifications and Training:

MBA/PGDM – Banking and Finance, NIBM Pune

B.Tech – IT, Government Engineering College, Thiruvananthapuram

Key Relevant Experience:

Validated PD, EAD, Recovery models for CCAR/DFAST/CECL

Conducted Annual Model Review, evaluated model assumptions

Reviewed credit rating, behavioral models, scorecards

Led CCAR/CECL model backtesting

Participated in automation initiatives using Python

Technical Skills:

CCAR/DFAST, PPNR, ICAAP, CECL, IFRS9

Behavioral models and scorecards

Python, SAS, SQL

----- Priyajit Bishayee svbfcb5.txt -----

Priyajit Bishayee

Role: Senior

Summary:

Priyajit is a Senior in EY’s Financial Services Risk Management practice. He has 8+ years of experience in Banking and Financial Services.

Qualifications and Training:

MSc Statistics, IIT Kanpur (2014–2016)

BSc Statistics, St. Xavier’s College, University of Calcutta (2011–2014)

Key Relevant Experience:

Validated $10B energy sector scorecard

Basel capital model validations (PD, EAD, LGD)

Operational Risk: Monte Carlo simulations, $30B RWA impact

Stress testing models for CCAR

ML/AI model validations (Logistic, Gradient Boosting)

Built account-level survival models, EAD trees

Developed CECL, IFRS9 models

Estimated Weighted Average Life (WAL) for CCAR reports

Mentored interns, drove audit resolution

Technical Skills:

Regulatory: Basel II/III, CCAR, IFRS9

Tools: R, SAS, Python, Excel

ML: Random Forest, Gradient Boosting, SVM, CART

Modeling: GLM, spline, logistic regression

Documentation: LaTeX, MS Word

----- Rahul Kushwaha svbfcb5.txt -----

Rahul Kushwaha

Role: Senior

Summary:

Rahul Kushwaha is a Senior in EY's Financial Services Risk Management practice with 2+ years of experience.

Worked on credit risk model validation, time-series analytics, and financial decision-making models.

Qualifications and Training:

MSc. Quantitative Economics, ISI New Delhi, 2023

B.A. Hons (Economics), University of Delhi, 2020

Key Relevant Experience:

Time series model validation for Securities, IB, Lending

Created interactive validation dashboards in Python

Designed loan disbursement analytics leading to ₹30Cr+ increase

Technical Skills:

Python, R, SQL, STATA, LaTeX, MS Office

----- Rubika Rana svbfcb4.txt -----

Rubika Rana – Senior Manager

Summary:

10+ years of experience in model development, monitoring, and validation of credit risk models.

Expertise in predictive analytics, CCAR & PPNR models, with model governance exposure.

Qualifications:

B.Tech, Thapar University, 2013

Key Experience:

Led development of PPNR & Balance Sheet models for CCAR.

Developed CCAR EWS models using Python (K-means clustering, trend analysis).

Scenario-based testing (Base, Adverse, Severely Adverse).

Used metrics: KS, Gini, ROC-AUC, PSI, MAE, RMSE.

Automated CCAR pipelines using Python.

Created interactive Python dashboards.

Complied with SR 11-7 and CCAR expectations.

Collaborated with Risk, Treasury, Finance teams.

Technical Skills:

IRB/CCAR model development and validation

SAS EG, SQL, Python

----- Rubika Rana svbfcb5.txt -----

Rubika Rana

Role: Senior Manager

Summary:

Rubika is a Senior Manager in EY. She has 10+ years of experience in model development, monitoring and validation of credit risk models.

She is proficient in predictive analytics, statistical modelling and has good exposure to CCAR & PPNR models, along with model development and validation of credit risk models.

Qualifications and Training:

B.Tech from Thapar University, 2013

Key Relevant Experience:

Model Development and validation experience:

Led development of PPNR and Balance Sheet models for annual CCAR submissions, including data preparation, variable selection, and multi-scenario calibration across Base, Adverse, and Severely Adverse scenarios.

Developed models using Linear Regression, Time Series Regression (ARIMA), and panel data techniques in R and Python to predict revenues across Investment Banking business units.

Developed CCAR EWS model by validating a suite of Early Warning Indicators (EWIs) to monitor credit deterioration under stress scenarios.

Combined rating migration analysis, portfolio clustering (K-means), and trend analysis using Python to enhance risk detection frameworks within CCAR models.

Conducted back-testing, sensitivity analysis, and scenario-based performance assessments for PPNR, Balance Sheet, and PD models used in CCAR submissions.

Validated forecast accuracy, stability under stress, and macroeconomic sensitivity using historical performance data and supervisory scenarios.

Utilized KS, Gini, ROC-AUC, and PSI to assess model discriminatory power, ranking performance, and population stability across CCAR baseline and stressed conditions to undertake model performance testing.

Developed interactive Python dashboards to track CCAR model performance, visualize key validation test results, and monitor PD migrations, revenue forecast deviations, and realized vs. projected outcomes.

Developed Python pipelines to automate CCAR model execution.

Authored model development documentation, validation reports, and performance summaries.

Addressed regulatory feedback per SR 11-7 and CCAR expectations.

Conducted workshops with Risk, Treasury, Finance, and Model Governance teams.

Supported regulatory exams and walkthroughs.

Technical Skills:

IRB/CCAR Credit Risk Model Development, Monitoring and Validation

Model risk management, model governance

Deep understanding of statistical modelling techniques

SAS EG, SQL, and Python

----- RubikaRana-anzey.txt -----

Senior Manager: Rubika Rana

Summary

Rubika works as a Senior Manager in the Financial Services Risk Management practice. She has 10+ years of experience in Credit risk and data science. Seasoned Credit Risk Analyst, specializing in model validation and optimization, skilled in data analysis, quality assurance, and predictive modelling for robust credit risk assessment and management.

Qualifications and training

Bachelors of Engineering, Thapar University, 2013

Key relevant experience

Mentored new joiners as vertical expanded from 1 to 50-member team; led a 5-member team.

Validated OCR models (RNN, LSTM, NLP) for Global Technology & Ops.

Validated high-risk models using Linear & Time Series Regression (Market Risk, PPNR).

Validated AML models using quantitative methods.

Led Decision Science Team for monitoring Credit Risk & Regulatory models (Application, Behavioural, Basel/IFRS9 – PD, EAD, LGD), repayment models – for S.A and U.K banks.

Developed K-Means clustering algorithms for:

Mobile/online developers (churn prevention)

Seller/client zone clustering

Built application & behavioural models using logistic regression to reduce portfolio default rate.

Created model for order prediction (rider allocation based on traffic).

Built client performance adjustment system based on weight factor.

Led IRB PD model development (DQ analysis, variable treatment, segmentation, calibration, MoC).

Assessed regulatory compliance of IRB models to CRDIV; prepared compliance plan.

Developed PPNR & Balance Sheet models for CCAR submission under various market scenarios.

Modeled Global Rates Trading revenue under IFRS9.

Realigned business strategies based on model performance insights.

Technical Skills

R, Python, SQL, SAS EG

IRB/CCAR Model Development, Monitoring, Validation

Model risk management, governance

AI/ML modelling, MS Office

Statistical modelling techniques

----- Sahil Patil svbfcb5.txt -----

Sahil Patil

Role: Senior

Summary:

Sahil is a Senior in EY’s FSRM practice with ~3 years of experience in credit risk model validation, ICAAP, and stress testing.

Qualifications and Training:

FRM Certified (GARP), 2024

MBA (Decision Sciences & Analytics), NMIMS, 2021

B.Com, R.A. Podar College, 2018

Key Relevant Experience:

Revalidated PD, EAD, LGD models for CCAR

Built delinquency bucket model using Python for RWA

Built stress testing, scenario analysis frameworks

Developed models for liquidity, credit, interest rate risk

Validated models via R-Ratio, Stress Ratio

Developed PD/LGD models per Ind AS 109 / IFRS 9

Drafted risk appetite statements

Built VAR model for trading book

Technical Skills:

Python, SAS, SQL, Power BI, Tableau

Linear/Logistic Regression, Decision Trees, Survival Analysis

IFRS9, CCAR, CECL, ICAAP, IRB, SR 11-7/15-18/15-19

----- Sakeena Siddiqui svbfcb1.txt -----

2. Sakeena Siddiqui — Senior

Summary

Sakeena is a Senior in EY’s Financial Services Risk Management practice. She has 5+ years of experience in the Banking and Financial Services industry. Proficient in CCAR/DFAST PD/LGD/EAD model development, and stress testing as per CCAR/CECL norms for Commercial portfolio.

Qualifications and training

B.Tech, Computer Science and Engineering, GGSIPU, Delhi

Key relevant experience

Implemented wholesale credit risk models for CCAR.

Authored model documentation for CCAR audits.

Refined PPNR estimates under macroeconomic stress.

Designed economic scenarios for stress testing (UK G-SIB).

Developed IFRS9 PD scorecard for personal exposure portfolio (Middle East Central Bank).

Built PD scorecards for Large Corporate & ABL portfolios for US banks under CECL.

End-to-end model development: data prep, segmentation, macro impacts, backtesting.

Validated models across retail/commercial, aligned with CCAR.

Highlighted data quality issues and recommended MRM resolution strategies.

Technical Skills

Basel II/III, CCAR, IFRS9, Credit Risk Models, Stress Testing, Capital Planning, Risk Analytics, Excel, SAS, Python, Model Risk Management, Model Audit

----- Saumya Bhardwaj hsbcifrs9cvs.txt -----

Saumya Bhardwaj

M.Sc. in Economics | GCB 5

Summary

Saumya is part of EY FSRM with 4+ years of experience. Her expertise lies in credit risk modelling, risk analytics, and Expected Loss Computation under IFRS9, with knowledge of SR11/7 and SS 1/23.

Experience

HSBC:

Developed IFRS9 model suite for SME portfolio

Performed data quality checks, variable selection, segmentation, model testing

LGD model development using SAS and Python

Sensitivity, stability analysis (bootstrapping, k-fold cross-validation)

Forecast model for PD projections and IFRS9 ECL

Model Validation:

Digital scorecard validation for Indian bank (Auto Loans)

ML scorecard models (GBM, XGBoost) for US G-SIB

Capital risk model validation for Principal Investments (FV, RWA, PVA under stress)

CLI/CLD impairment and credit strategy model validation (UK G-SIB)

----- Sautrik Ganguly svbfcb5.txt -----

Sautrik Ganguly

Role: Senior

Summary:

Sautrik is a Senior with 2 years of experience in CCAR model development and execution. Specialized in PD, CCR, and stress testing.

Qualifications and Training:

MA Economics, Delhi School of Economics (2023)

BSc Statistics, RKMRC Narendrapur (2020)

Key Relevant Experience:

Developed CCAR models (PD, Loss Aggregation)

Conducted CCR monitoring using Monte Carlo simulation

Validated risk models (BT Seaborg, Credit Limit Decrease)

Created Card Risk Group Segmentation Model

Built automation tools for model monitoring in Python

Designed predictive stock clearance models

Technical Skills:

Python (certified), STATA

MS Excel, Word

CCAR, CCR model validation

Candidate 16: Bhawya Mehra

Role: Senior

Summary:

Bhawya is a Senior in EY’s Financial Services Risk Management practice. She has more than 3 years of experience in model development, performance monitoring, and financial analytics, specializing in CCAR/DFAST modeling, stress testing, and economic data-driven forecasting.

Qualifications and Training:

MA Economics, Madras School of Economics, Delhi (2019 – 2021)

BA (H) Economics, University of Delhi (2016 – 2019)

Key Relevant Experience:

Developed and implemented a consolidated PPNR modeling framework for a leading US Bank, improving forecasting accuracy for CCAR submissions and business planning.

Integrated macroeconomic indicators and strategic business drivers (ECM, M&A pipeline data) to optimize revenue prediction models.

Redeveloped CCAR FX models for ICG Markets portfolios, designing OLS-based econometric models using Python/SAS to project total revenue and client revenue based on macroeconomic variables such as currency volatility index.

Applied advanced econometric and statistical methodologies, including time-series analysis, residual diagnostics, seasonality adjustments, outlier detection, and regression with dummy variables, to capture portfolio sensitivities to economic shocks (COVID-19, geopolitical crises).

Collaborated with Model Validation teams to achieve approval for champion models, ensuring alignment with regulatory expectations and stress-testing requirements.

Pioneered a novel methodology to adjust revenue projections in response to market shifts, integrating real-time business intelligence and strategic adjustments.

Led ongoing model performance monitoring, conducting back-testing, sensitivity analysis, and stability testing to assess model robustness and identify potential breaches or limitations flagged by Model Risk Management (MRM).

Authored comprehensive model documentation including Model Development, Recalibration Reports, Performance Monitoring, Limitation Record Change, and Revalidation Reports, ensuring compliance with regulatory and audit requirements.

Developed a predictive model for loan delinquency for a UK-based G-SIB, leveraging machine learning algorithms and advanced data treatment techniques such as outlier analysis, association analysis, oversampling, and feature selection to improve model accuracy.

Implemented data visualization and statistical analysis to identify key drivers of loan defaults, enhancing risk assessment and decision-making for credit approvals.

Technical Skills:

Experience of regulatory regimes viz. CCAR, Stress testing models

Model development, monitoring, visualization

Python, R, SQL, Advanced Excel

----- Senior Kshitij Sahdev anzscorecards.txt -----

Senior: Kshitij Sahdev

Summary:

Kshitij is a Senior within EY's Financial Services Risk Management practice, with over 3 years of specialized experience in the Banking and Financial Services sector. Proficient in leveraging advanced machine learning techniques for data analysis and predictive modelling. He is proficient with SAS, Python and Pyspark.

Qualifications and training:

Masters in Environmental Economics, Madras School of Economics, 2021

Bachelors in Economics, Delhi University, 2018

Key relevant experience:

Developed PD model for retail portfolio (UK bank) using XGBoost

Validated PD model for US bank retail portfolio

Created Geographic/Industry scores for business health

Automated Country Risk/MDA scores using PCA & RapidFuzz

Converted SAS scorecard codes to PySpark

Performed multivariate time-series forecasting using SARIMAX

Estimated economic impact on workforce and GDP

Analyzed maritime shipment delays and bridge collapse effects

Used Denoising Autoencoders (MIDAS) for missing value imputation

Applied HuggingFace LLMs for text summarization and sentiment analysis

Created Covid-19 Impact Index

Produced monthly investigative reports on identity theft/misrepresentation

Technical Skills:

Predictive modelling, Scorecards, Machine Learning

SAS, Python, Pyspark

Deep Learning, Deep statistical modelling

----- Senior Stuti Mehrotra anzscorecards.txt -----

Senior: Stuti Mehrotra

Summary:

Stuti is a Senior within EY's Financial Services Risk Management practice, with over 2 years of specialized experience in the Banking and Financial Services sector. Proficient in the development and validation of scorecard models, PPNR models within the CCAR framework. She is proficient with SAS, Python, Tableau.

Qualifications and training:

M.Sc. Financial Economics, Gokhale Institute of Politics and Economics, 2022

BA Honors Economics, University of Delhi, 2018

Key relevant experience:

Validated retail scorecard model for US bank

Developed PD scorecard model for UK bank

Developed and implemented PPNR models for CCAR 9Q forecasting

Conducted stress testing for model resilience

Developed ongoing monitoring framework

Used multivariate linear regression on time-series data

Designed Tableau dashboards to visualize CCAR outputs

Resolved data inconsistencies in revenue forecasting

Technical Skills:

Scorecard Model Development/Validation, PPNR, Stress Testing, Risk Analytics

SAS, Python, R & SQL, Advanced Excel, Tableau

Model Risk Management, Regulatory Regimes (Basel II/III, CCAR, IFRS9)

----- Senior Vedanti Khokher anzscorecards.txt -----

Senior: Vedanti Khokher

Summary:

Vedanti is a Senior within EY's Financial Services Risk Management practice, with close to 3 years of specialized experience in the Banking and Financial Services sector. Proficient in the end-to-end development of scorecard models using SAS and credit risk analytics. She is proficient with SAS, SQL and Python.

Qualifications and training:

B.Tech in Chemical Engineering, IIT Roorkee, 2022

Key relevant experience:

Developed PD scorecard model for UK bank (Singapore and Malaysia market)

Led logistic regression scorecards across 7 customer segments

Achieved AR lift of ~8%, resolved PSI/CSI breaches

Validated APAC bank’s retail model and built challenger

Designed approve-decline strategies for Hong Kong credit card portfolio

Assigned SMV risk grades based on credit scores

Logistic regression model identified riskiest 50% population with 2.5x default rate

Technical Skills:

SAS, SQL, Python, Excel

Scorecard Modelling, Credit Risk Analytics, Credit Decisioning

----- Shambhu Dayal Singh svbfcb1.txt -----

1. Shambhu Dayal Singh — Senior Manager

Summary

Shambhu is a Senior Manager in EY. He has 9+ years of experience in financial domain with exposure across model development, monitoring and validation of credit risk models. He is proficient in predictive analytics, statistical modelling and has good understanding of various modelling techniques.

Qualifications and training

M.Tech from Indian Statistical Institute, Kolkata, 2015

B. Tech. from NIFFT, Ranchi, 2011

Key relevant experience

Led development of economic scenarios for CCAR stress testing.

Integrated macroeconomic variables into credit risk models.

Oversaw data management process for CCAR models.

Contributed to credit Risk (PD & LGD) Models Fit-For-Use Analysis under B3R for a Canadian Bank.

Calculated RWA impact due to population extension (~21% reduction).

Analyzed impairments for BCUS portfolio; supported ALLL and IFRS9 analytics for US region.

Developed Application Scorecard and PD model (IFRS9) for digital loans in India.

Re-developed Behavioural Scorecard for UK loans portfolio.

Credit-risk impairment forecasting via regression and balance-ratio models.

Converted IAS 39 forecasting templates to IFRS 9.

Trigger review, model validation for PD model for Canadian Bank (KS drop, IV analysis).

Technical Skills

Basel II/III, CCAR, IFRS9, Credit Risk Models, Stress Testing, Capital Planning, Risk Analytics, Excel, SAS, Python, Model Risk Management, Model Audit

----- Shashwata Mondal anzey.txt -----

Shashwata Mondal

Summary

Shashwata is a Senior in EY’s Financial Services Risk Management practice. He has 1+ years of experience in the Banking and Financial Services industry. Experienced Credit Risk Analyst with a strong background in predictive modelling adept at validating and refitting credit models, with hands-on expertise in various ML techniques.

Qualifications and training

M.Tech in Cryptology and Security, ISI Kolkata, 2023

M.Sc in Mathematics, Ramakrishna Mission Vivekananda Educational and Research Institute, 2020

B.Sc in Mathematics, Ramakrishna Mission Residential College, 2018

Key relevant experience

Data preprocessing (outlier handling, missing values, scaling, encoding)

Used XGBoost (classification: claim/no-claim; regression: claim amount)

Tuned hyperparameters, validated with K-Fold Cross-Validation

Evaluated using metrics like Confusion Matrix, R², TPR, Adjusted R²

Logistic Regression for PD modelling; Linear Regression for LGD

Computed Expected Loss using PD, LGD, fixed EAD

Feature mapping for marine insurance: PSL, sum insured, claims, commodities

Used internal calculator to estimate expected premium

Compared expected vs actual premium rates

Built system for month-on-month zonal rate computation

Monitored geographic rate trends for pricing strategy

Technical Skills

R, Python, SQL, C, Git, Excel

Machine Learning, AI

Statistics, Deep Learning, NLP, Time Series

----- Shashwata Mondal svbfcb5.txt -----

Shashwata Mondal

Role: Senior

Summary:

Shashwata is a Senior in EY’s Financial Services Risk Management practice. He has close to 2 years of experience in the Banking and Financial Services industry.

Qualifications and Training:

M.Tech in Cryptology and Security, ISI Kolkata, 2023

M.Sc in Mathematics, RKM Vivekananda University, 2020

B.Sc in Mathematics, RKM Residential College, 2018

Key Relevant Experience:

Executed CCAR/CECL stress testing and forecasting

Delivered outputs for 9 cohort recovery and GCLR models (CCAR 2020)

Developed scorecards using XGBoost (classification + regression), validated using K-Fold

Modeled PD (logistic regression), LGD (linear regression)

Applied feature scaling, outlier handling, encoding for preprocessing

Computed Expected Loss with PD, LGD, fixed EAD

Analyzed expected premium vs actual premium (marine insurance)

Created monthly zonal rate monitoring systems

Technical Skills:

Python, R, SQL, C, Git, Excel

ML, AI, Deep Learning, NLP, Time Series

----- Shriwari Mhatre hsbcifrs9cvs.txt -----

Shriwari Mhatre

MSc in Data Science and ML | GCB 6

Summary

Shriwari is part of EY FSRM India with 5+ years’ experience in credit risk management, ML model development/validation across Banking and FMCG.

Experience

Model Validation (AMEX):

Validated Credit Risk and Fraud models (US + global portfolios)

Techniques: logistic regression, time series, fractional regression

Performed OOT/OOS testing, backtesting, sensitivity/stress testing, calibration

Drafted observations and recommendations

Other:

Developed B-score for credit card cross-sell

Built acquisition scorecards using in-house + alt data

Demand forecasting/replenishment model (80% accuracy, $15M impact)

MLOps solution hosted on Azure with Databricks

Deloitte internal supply chain asset – 35M+ opportunities onboarded

----- Sneha Khanna hsbcimr.txt -----

Sneha Khanna

MSc in Economics

GCB 6

Summary

Sneha is a Senior in EY FSRM India with 4+ years of experience. She has supported global banks on Liquidity, Credit, and Climate Risk models.

Experience

Model Validation

Validated high-materiality models: Mortgage Pipeline, Wholesale Funding

IFRS 9 PD & Retail Loan models: Transition Matrix, Pluto Tasche, Linear Regression

Sensitivity analysis (parameter/input)

Reviewed RRP model (legal frameworks, development/validation docs)

Built PA scoring model (water vehicle portfolio, SAS)

Other Experience

Automated Testing for Unsecured Credit Model (SAS, Excel, SoapUI)

Python-based Automated Data Extraction

Data quality checks, stakeholder data gathering

Built Climate Sector GVA models (Python, R)

----- Stuti Mehrotra svbfcb3.txt -----

Stuti Mehrotra

Role: Senior

Summary:

Stuti is a seasoned professional in EY's Financial Services Risk Management practice with over 2 years of specialized experience in the Banking and Financial Services sector.

Proficient in developing and validating PPNR forecasting models for capital planning and revenue projections under the CCAR framework.

Qualifications and Training:

M.Sc. Financial Economics, Gokhale Institute of Politics and Economics, 2022

B.A. (Hons) Economics, University of Delhi, 2018

Technical Skills:

Regulatory experience: Basel II/III, CCAR, IFRS9

Tools: SAS, Python, R, SQL, Advanced Excel, Tableau

Expertise: Model Development, Validation, Stress Testing, Risk Analytics, Model Risk Management

Key Relevant Experience:

Independently developed and implemented PPNR models for CCAR 9Q forecasting at a major US financial institution.

Designed and executed stress testing to assess model resilience under adverse economic conditions.

Created an Ongoing Monitoring framework for tracking model performance.

Applied advanced econometrics (e.g., Multivariate Linear Regression on time-series data) to forecast financial metrics.

Collaborated cross-functionally with risk, finance, and IT for strategic model integration.

Validated models via back-testing, scenario/sensitivity analysis, and coefficient stability checks.

Cleaned and verified financial datasets used for CCAR reporting and capital stress testing.

Designed a Tableau dashboard to visualize CCAR model outputs for improved stakeholder understanding.

----- Stuti Mehrotra svbfcb4.txt -----

5. Stuti Mehrotra – Senior

Summary:

2+ years in BFSI, skilled in PPNR model development and validation for CCAR.

Qualifications:

M.Sc. Financial Economics, Gokhale Institute (2022)

BA Economics, Delhi University (2018)

Key Experience:

Led PPNR model development and validation (9Q CCAR forecasts).

Stress testing protocols and ongoing monitoring.

Multivariate regression forecasting for key financials.

Inter-department collaboration with risk and finance.

Developed Tableau dashboards for CCAR outputs.

Technical Skills:

SAS, Python, R, SQL, Excel, Tableau

Basel II/III, CCAR, IFRS9

Risk analytics, stress testing

----- Stuti Mehrotra svbfcb5.txt -----

Stuti Mehrotra

Role: Senior

Summary:

Stuti is a Senior in EY’s FSRM practice with 2+ years of experience, specializing in PPNR and CCAR.

Qualifications and Training:

M.Sc Financial Economics, Gokhale Institute, 2022

BA (H) Economics, Delhi University, 2018

Key Relevant Experience:

Developed CCAR PPNR models (9Q forecasting)

Designed stress testing protocols

Built Ongoing Monitoring framework

Applied MLR on Time Series for issuance, income, opex

Collaborated across risk, finance, IT

Conducted back-testing, coefficient stability checks

Verified datasets for regulatory submissions

Built Tableau dashboards to visualize PPNR outputs

Technical Skills:

Basel II/III, CCAR, IFRS9

SAS, Python, R, SQL

Advanced Excel, Tableau

----- Subhadip Acharyya svbfcb5.txt -----

Subhadip Acharyya

Role: Senior

Summary:

Subhadip has over 10 years of experience in model development, monitoring, and validation. Expert in stress testing, scenario generation, and market risk.

Qualifications and Training:

MSc Statistics, IIT Kanpur

BSc Statistics, Presidency College, CU

Key Relevant Experience:

Scenario expansion modeling for CCAR, LPA

Built time-series models for asset class forecasts

Closed NFRI, audit/validation issues

Delivered LIBOR transition, climate risk models

Owned end-to-end model lifecycle, including governance

Developed scenario path models (daily price shocks)

Built Value at Risk (VaR) and PD models for US portfolios

Technical Skills:

Credit Risk, PPNR, IRB Modeling

Time Series, Econometrics, Market Risk

Tools: R, SAS, Python

Strong in documentation and project management

----- Sulagna Paul svbfcb5.txt -----

Sulagna Paul

Role: Senior Manager

Summary:

Sulagna is a Senior Manager with 9+ years of experience in credit risk. Expert in CECL, CCAR, IFRS9, and Basel models.

Qualifications and Training:

Masters in Economics, Calcutta University (2014)

B.Sc (H) Economics, Asutosh College (2012)

Key Relevant Experience:

Developed & validated CCAR/CECL/IFRS9 models

Built State Transition Models (Logistic, Linear Regression)

Validated stress testing models under extreme scenarios

Reviewed FRB-compliant secured portfolio models

Built PD model for UK mortgage (approved by PRA)

Led behavioral credit risk models using clustering/factor analysis

Drove process automation and mentoring

Technical Skills:

CCAR, IFRS9, Basel

Base SAS, SAS E-miner, R, Excel

----- Susmita Misra svbfcb2.txt -----

Susmita Misra

Summary

Susmita is a Senior in the Financial Services Risk Management practice of EY India focusing on Credit Risk Analytics, regulatory modelling, Model monitoring and validation with 3 years of experience. She has worked largely with credit model development and validation team for Global Banks in financial services domain.

Qualifications and training

B.Sc. in Statistics from Guwahati University in 2018

Post Graduation Diploma in Statistics and Analytics from ISI in 2019

Key relevant experience

For a European Bank, worked with the ERM PPNR Stress testing team.

Responsible for the syndication of PPNR-stressed projections for the IB division.

Quantified and assessed the capital requirements for business risk under semi-annual and quarterly internal group stress testing exercises.

Contributed to the design and calibration of econometric models for PPNR forecasting, focusing on deposit and fee income portfolios.

Supported the execution of CCAR scenario analysis for trading and asset management portfolios, providing insights into market and operational risks.

Managed the PPNR modelling process for consumer credit portfolios, incorporating behavioural and macroeconomic variables to improve forecast accuracy.

Coordinated with risk management and finance teams to integrate CCAR stress testing results into the bank's capital planning and risk appetite framework.

Played a pivotal role in the end-to-end process of CECL model development, from data collection to model deployment, for mortgage and commercial loan portfolios.

For a G-SIB bank, validated their logistic regression-based PD models for unsecured Credit portfolio.

Conducted back-testing, benchmarking, and sensitivity analysis of the CCAR models for the bank’s quarterly and annual reviews.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit

----- Swar Sudha hsbc.txt -----

Swar Sudha

MBA

GCB 6

Skill

IFRS9 ✓ SAS/SQL/Python ✓

Model Validation ✓ Credit Scorecards ✓

Machine Learning IRB

Summary

Swar is a part of the Financial Services Risk Management practice of EY India with 2+ years of professional experience in Credit Risk and IRB.

Has professional experience working on credit risk, IRB, scorecard models and based norms.

Experience

Model Validation

Conducted regular reviews—quarterly and annually—of Basel IRB credit and scorecard models. This involved identifying performance trends, anomalies, and ensuring models remained accurate and up-to-date.

Performed in-depth qualitative and quantitative analysis on IRB models. This analysis aimed to validate model integrity, robustness, and accuracy, ensuring they met both regulatory and internal standards.

Evaluated the performance of models by comparing them against development standards and industry benchmarks. This process helped in identifying areas for improvement and ensuring models met expected performance levels.

Other Experience

Collaborated with various regulators and adhered to regulatory frameworks such as Basel II & III.

Prepared comprehensive performance reports and presentations. These were shared with stakeholders, providing them with clear insights into model performance and any required actions or improvements.

Utilized advanced tools like SAS, SQL, and Python to conduct rigorous quantitative testing. This involved running simulations, performing data analysis, and validating model outputs to ensure accuracy and reliability.

Documented the review and challenge process, providing clear rationales for each threshold decision. This documentation ensured transparency and accountability in the threshold-setting process.

Maintained regular communication with project leadership, providing updates and progress reports. This ensured all stakeholders were informed and any issues were promptly addressed.

----- Tarun Anand svbfcb1.txt -----

4. Tarun Anand — Consultant

Summary

Tarun is a consultant in the Financial Services Risk Management practice at EY India with 1+ years of experience. His expertise covers a range of complex areas, which include model development, policy creation, model validation, and monitoring of credit risk models. He has worked with banks and NBFCs across India, Southeast Asia, and the Middle East.

Qualifications and training

BA(H) Economics, Loyola College

MA Economics, Madras School of Economics

Technical Skills

Basel II/III, CCAR, IFRS9, Credit Risk Models, Stress Testing, Capital Planning, Risk Analytics, Excel, SAS, Python, Model Risk Management, Model Audit

----- Vaibhav Yadav svbfcb2.txt -----

Senior – Vaibhav Yadav

Summary

Vaibhav is a Senior in EY’s Financial Services Risk Management practice. He has 4+ years of experience in the Banking and Financial Services industry. Proficient in CCAR/DFAST PD/LGD/EAD model development, and stress testing as per CCAR/CECL norms for Commercial portfolio.

Qualifications and training

MA Economics, Delhi University, 2021

BA(H) Economics, Delhi University, 2019

Key relevant experience

Contributed to Macro Modelling and model monitoring for Interest Rates and Macro variables.

Quarterly reviewed and adjusted projection paths for these variables under different scenarios including CCAR BHC & FRB, 5 Firm internal Macro Scenarios, and Climate FRBCSA pilot project scenarios.

Worked on Climate Fed Pilot Project to create long term (40Q) forecasts for variables, along with Modelling granular sector level GVAs to better capture Climate scenario impact.

Owned Geopolitical Asia Escalation scenario with in Firm Macro Scenarios and delivered quarterly forecasts.

Worked on dataset production for ~1300 user claimed variables.

Engaged in DQ checks, usual data fixes, sourcing data from multiple data providers, running scenarios and creating projection sets.

Worked on python script to make the DQ checks process more focused by significantly detecting 98% (350k Standalone Issues and 700k comparison issues) as non-material and recurring and introduced priority segmentation of issues significantly reducing the effort.

Engaged in quantitative model validation of PPNR models in accordance with CCAR and SR 11-7 Guidelines.

Analysed model methodology and formulation; risk associated with model assumptions and limitations.

Appraising the input data; conducting statistical testing of variables and model residuals using SAS.

Evaluating model outputs; overrides and overlays.

Assisted in the validation and back-testing of CECL models for consumer lending portfolios, ensuring compliance with accounting standards and regulatory expectations.

Led monthly macro monitoring, prepared deck and Flash Reports (presented to downstream users) regarding recent developments in economic environment.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit

----- Vaibhav Yadav svbfcb5.txt -----

Vaibhav Yadav

Role: Senior

Summary:

Vaibhav is a Senior in EY’s FSRM practice with 4+ years of BFSI experience. Skilled in CCAR, macro-modelling, data lineage.

Qualifications and Training:

MA Economics, Delhi University, 2021

BA (H) Economics, Delhi University, 2019

Key Relevant Experience:

Reviewed macro projections under CCAR BHC & FRB

Validated PPNR models under CCAR, SR 11-7

Validated CECL models for consumer lending

Built long-term forecasts for Climate Fed Pilot

Created geopolitical macro scenarios

Handled data production (~1300 variables)

Automated DQ checks (reduced false flags by 98%)

Led macro monitoring and stakeholder flash reports

Technical Skills:

Basel II/III, CCAR, IFRS9

Advanced Excel, SAS, Python

Model Risk Management, Capital Planning

----- Yash Rai svbfcb2.txt -----

Yash Rai

Summary

Yash is a Senior in EY’s Financial Services Risk Management practice. He has 4 years of experience in data analytics projects, collaborating with cross functional teams and ensuring the accuracy and integrity of data and actionable insight. Skilled in statistical modelling, data analysis, predictive modelling, text mining, machine learning and deep learning.

Qualifications and training

B.E, Chitkara University

Key relevant experience

Applied SAS to construct and calibrate CCAR models, simulating balance sheet and P&L impacts under various stress scenarios to inform capital planning decisions for mid-size commercial loan portfolios.

Designed SQL queries to retrieve and cleanse financial data, ensuring the integrity and consistency of inputs used in PPNR models for accurate revenue forecasting.

Refined PPNR models using Python to incorporate new product features and changing business strategies, ensuring models remained relevant and aligned with market conditions.

Conducted robust back-testing and sensitivity analysis for CCAR models using SAS, identifying key risk drivers and enhancing the predictive power of capital stress tests.

Implemented CECL models in compliance with accounting standards, leveraging Python to integrate historical loss data with forward-looking economic indicators for consumer loan portfolios.

Enhanced PPNR forecasting models with advanced econometric techniques in SAS, capturing the nuances of non-interest income and expense streams for asset management divisions.

Collaborated with risk management teams to align CCAR model outputs with enterprise-wide risk appetite frameworks, providing critical input for risk mitigation strategies.

Developed comprehensive documentation for CECL model methodologies, assumptions, and validation results, facilitating clear communication with auditors and regulatory examiners.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit

----- svbfcb2.txt -----

Senior – Vaibhav Yadav

Summary

Vaibhav is a Senior in EY’s Financial Services Risk Management practice. He has 4+ years of experience in the Banking and Financial Services industry. Proficient in CCAR/DFAST PD/LGD/EAD model development, and stress testing as per CCAR/CECL norms for Commercial portfolio.

Qualifications and training

MA Economics, Delhi University, 2021

BA(H) Economics, Delhi University, 2019

Key relevant experience

Contributed to Macro Modelling and model monitoring for Interest Rates and Macro variables.

Quarterly reviewed and adjusted projection paths for these variables under different scenarios including CCAR BHC & FRB, 5 Firm internal Macro Scenarios, and Climate FRBCSA pilot project scenarios.

Worked on Climate Fed Pilot Project to create long term (40Q) forecasts for variables, along with Modelling granular sector level GVAs to better capture Climate scenario impact.

Owned Geopolitical Asia Escalation scenario with in Firm Macro Scenarios and delivered quarterly forecasts.

Worked on dataset production for ~1300 user claimed variables.

Engaged in DQ checks, usual data fixes, sourcing data from multiple data providers, running scenarios and creating projection sets.

Worked on python script to make the DQ checks process more focused by significantly detecting 98% (350k Standalone Issues and 700k comparison issues) as non-material and recurring and introduced priority segmentation of issues significantly reducing the effort.

Engaged in quantitative model validation of PPNR models in accordance with CCAR and SR 11-7 Guidelines.

Analysed model methodology and formulation; risk associated with model assumptions and limitations.

Appraising the input data; conducting statistical testing of variables and model residuals using SAS.

Evaluating model outputs; overrides and overlays.

Assisted in the validation and back-testing of CECL models for consumer lending portfolios, ensuring compliance with accounting standards and regulatory expectations.

Led monthly macro monitoring, prepared deck and Flash Reports (presented to downstream users) regarding recent developments in economic environment.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit

Senior – Susmita Misra

Summary

Susmita is a Senior in the Financial Services Risk Management practice of EY India focusing on Credit Risk Analytics, regulatory modelling, Model monitoring and validation with 3 years of experience. She has worked largely with credit model development and validation team for Global Banks in financial services domain.

Qualifications and training

B.Sc. in Statistics from Guwahati University in 2018

Post Graduation Diploma in Statistics and Analytics from ISI in 2019

Key relevant experience

For a European Bank, worked with the ERM PPNR Stress testing team.

Responsible for the syndication of PPNR-stressed projections for the IB division.

Quantified and assessed the capital requirements for business risk under semi-annual and quarterly internal group stress testing exercises.

Contributed to the design and calibration of econometric models for PPNR forecasting, focusing on deposit and fee income portfolios.

Supported the execution of CCAR scenario analysis for trading and asset management portfolios, providing insights into market and operational risks.

Managed the PPNR modelling process for consumer credit portfolios, incorporating behavioural and macroeconomic variables to improve forecast accuracy.

Coordinated with risk management and finance teams to integrate CCAR stress testing results into the bank's capital planning and risk appetite framework.

Played a pivotal role in the end-to-end process of CECL model development, from data collection to model deployment, for mortgage and commercial loan portfolios.

For a G-SIB bank, validated their logistic regression-based PD models for unsecured Credit portfolio.

Conducted back-testing, benchmarking, and sensitivity analysis of the CCAR models for the bank’s quarterly and annual reviews.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit

Senior – Yash Rai

Summary

Yash is a Senior in EY’s Financial Services Risk Management practice. He has 4 years of experience in data analytics projects, collaborating with cross functional teams and ensuring the accuracy and integrity of data and actionable insight. Skilled in statistical modelling, data analysis, predictive modelling, text mining, machine learning and deep learning.

Qualifications and training

B.E, Chitkara University

Key relevant experience

Applied SAS to construct and calibrate CCAR models, simulating balance sheet and P&L impacts under various stress scenarios to inform capital planning decisions for mid-size commercial loan portfolios.

Designed SQL queries to retrieve and cleanse financial data, ensuring the integrity and consistency of inputs used in PPNR models for accurate revenue forecasting.

Refined PPNR models using Python to incorporate new product features and changing business strategies, ensuring models remained relevant and aligned with market conditions.

Conducted robust back-testing and sensitivity analysis for CCAR models using SAS, identifying key risk drivers and enhancing the predictive power of capital stress tests.

Implemented CECL models in compliance with accounting standards, leveraging Python to integrate historical loss data with forward-looking economic indicators for consumer loan portfolios.

Enhanced PPNR forecasting models with advanced econometric techniques in SAS, capturing the nuances of non-interest income and expense streams for asset management divisions.

Collaborated with risk management teams to align CCAR model outputs with enterprise-wide risk appetite frameworks, providing critical input for risk mitigation strategies.

Developed comprehensive documentation for CECL model methodologies, assumptions, and validation results, facilitating clear communication with auditors and regulatory examiners.

Technical Skills

Experience of regulatory regimes viz. Basel II/III, CCAR, IFRS9

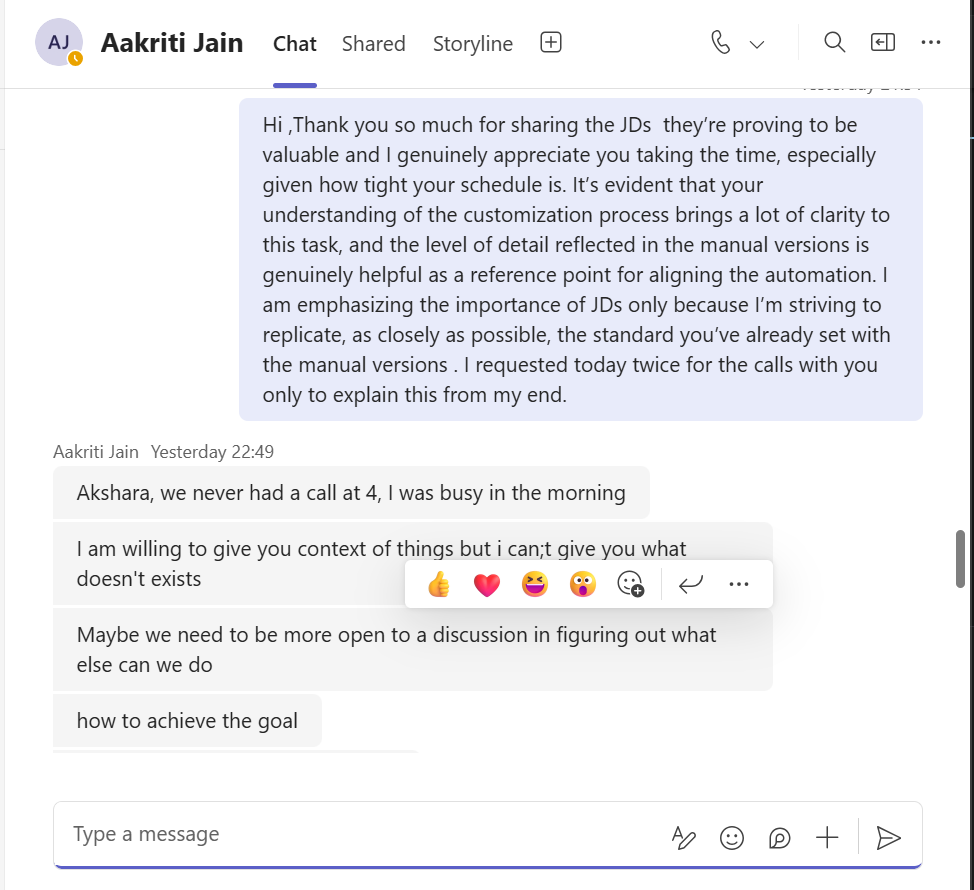
Model development and validation of credit risk models

Stress Testing, Capital Policy & Capital Planning, CCAR Risk Analytics

Advanced Excel, SAS, Python

Model Risk Management

Model Audit



A screenshot of a chat

AI-generated content may be incorrect.

A screenshot of a message

AI-generated content may be incorrect.

Hi [Senior Manager’s Name],

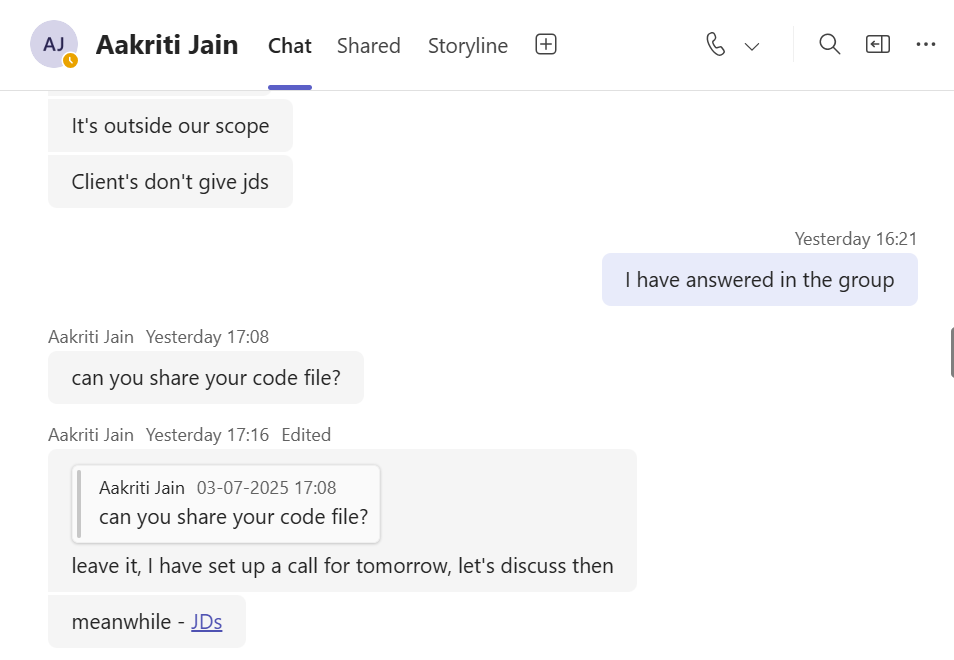
I understand you're currently out of office, and I truly apologize for reaching out during this time. I wouldn’t have done so unless I felt it was genuinely important to escalate this constructively and transparently.

Over the last few days, I’ve remained fully committed to aligning the LLM-based resume customization task with maximum precision. With the JDs folders finally shared, I am achieving match scores significantly improved — CV\_Aarsh (82.83%), Isha Porwal (80.58%), Kshitij (76.28%), and Kartik Mohan (77.63%) are now matching the sample resumes at over **70%+ accuracy**, which was not previously the case. This validates the **technical necessity of JD anchoring** to replicate the manual precision standard that’s already been set.

I’ve been with utmost sincerity willing to work collaboratively across the team - proactively requesting calls and overcoming technical blockers with apt solutions from my end . But I’ve faced resistance despite explaining the rationale clearly. I’m flagging this now only to ensure the outcome of the workstream stays intact and scalable.

Once you're back, I’m happy to walk you through both the technical process, results and blockers in more detail. Thank you for your time and understanding.

Warm regards,  
Akshara



Hi Sapta,

I am requesting few minutes of your time keep you informed with some context from my end. I’ve been making every effort to collaborate and align across stakeholders on the resume customization task. I’ve proactively initiated conversations — via direct messages, group chats, and emails with the sole intent of ensuring we’re on the same page about the JD alignment and execution path.

As the JDs were recently shared, and thanks to those, we’re now seeing encouraging outcomes. For instance, resumes like Aarsh (82.83%), Isha Porwal (80.58%), Kshitij (76.28%), and Kartik (77.63%) are all now showing **LLM-to-manual match scores exceeding 70%**, validating that these JDs are central to automation accuracy. Without the right JD anchoring, the scores drop significantly poor.

I continue to stay fully committed and open to working constructively with the team. I’m surfacing this now because I believe the current challenges aren’t stemming from a lack of willingness on my part but rather from misalignment or resistance I’ve been facing despite best efforts.

Warm regards,  
Akshara

Hi Nitin and Sapta,

I’ve successfully completed the scaling of the JD–Resume customization, with 29 resumes now tailored to their respective JDs and curated in the EY format. While the core work is in place, I’ve identified areas for content refinement and am currently working on further cleaning and enhancing the quality.

Dear Nimish

Hope you're doing well.

I wanted to share a quick summary of the workstream completed around the LLM-based resume customization process:

* Resumes with **well-structured content and clear experience mapping** achieved 75–83% match scores.
* Others with **vague, unstructured input** or limited alignment to the JD saw reduced effectiveness.
* Score variations confirmed the critical impact of **resume quality, JD relevance, and sample adherence** on final output.

Using the prompt variations defined in the attached Excel sheet given by Aakriti, I’m working on:

* Structuring **prompt-driven customization flows** (even without JD text).
* Experimenting with **prompt-category mapping** for roles and domains to ensure alignment even when explicit JDs are missing.
* Exploring improvements to formatting fidelity and experience restructuring.

Best regards,

The development is progressing well. I’ve shifted focus slightly toward automating the JD synthesis process using the shared prompts, since the tool needs to handle cases where no direct JDs are provided. I’ve generated 145 relevant JDs so far this step has taken around 356 minutes and will begin aligning resume customization with these next.

Just wanted to check in regarding work mode for the coming days. I’m progressing well on the JD-resume customisation task and currently working remotely. Do let me know if there's any change required based on project priorities.

Thanks, Nitin.Just checking,can I assume remote work will be okay for the rest of this week as well? I need to plan my travel and make ticket bookings accordingly. Will align based on your input.

Noted, thanks Nitin. I’ll proceed accordingly and keep you posted on the progress. Happy to connect whenever you're available to walk you through the task in detail.

Just wanted to keep you in the loop, I had a brief check-in with Nitin and I'm working remotely this week while progressing on the JD-resume customization task and will continue aligning based on project needs.\

**8/7/2025**

**✅ Goal**

Generate a *customized resume* for **each candidate** using **only the JDs whose skills semantically match** that candidate’s resume, and fill in EY's template.

**✅ Outline of Steps**

1. **Extract and read all files**:
   * JDs (from .txt files)
   * Raw resumes (.pdf/.docx)
   * EY format (.txt)
2. **Extract candidate skillset** from raw resumes (by extracting key skill-like words).
3. **Skill-match**: For each candidate, match their skills against the JD contents using LLM (Mistral 7B via LM Studio).
4. **Select matching JDs**: Keep only JDs with semantic similarity score ≥ threshold (e.g., 0.6).
5. **Customize resume** using the selected JD(s): Match candidate sentences to JD, and rephrase/match tone using LLM.
6. **Save output**: Save one final customized resume per candidate with their name and JD tag.

**✅ Assumptions**

* LM Studio is running at: http://localhost:1234
* Best available model: mistral-7b-instruct-v0.3-q4\_k\_m-gguf
* JD files are skill-based.
* You cannot use Sentence Transformers or external APIs.

Hi Aakriti,  
I’ve completed the JD synthesis (~145 prompts refined over 6–7 hours), and the resume customization code run is actively progressing—it’s compute-intensive and expected to complete by Thursday end. I’ll share outputs once ready.

**OLD CODE-2**

When an enduser uses the code through an application and gives any prompt , code to work from backend refers to data.Data here is defined set of prompts historically used and stored as Database

Here’s a **step-by-step explanation** of what your code does, broken down into clear sections:

**🔧 1. Imports and Libraries**

python

CopyEdit

import os, zipfile, fitz, docx, re, json, time, requests

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.metrics.pairwise import cosine\_similarity

* **File & folder handling**: os, zipfile
* **Reading resumes**: fitz (PDF), docx (Word), open (text)
* **Regex parsing**: re
* **Interacting with LM Studio**: requests, json
* **Model similarity**: TfidfVectorizer, cosine\_similarity

**📁 2. Configuration**

python

CopyEdit

base\_path = ...

Defines your base folder and sets paths to:

* Rawresumes.zip
* JDs.zip
* EY\_sample\_resume\_template.txt
* Folders: rawresumes, JDs, final\_customised\_resumes

Creates those folders if they don't already exist:

python

CopyEdit

os.makedirs(folder, exist\_ok=True)

**🧰 3. Utility Functions**

**3.1 Extract ZIP files**

python

CopyEdit

extract\_zip(zip\_path, extract\_to)

Extracts resume and JD ZIP files to folders.

**3.2 Read PDF or DOCX or TXT**

python

CopyEdit

read\_pdf\_text, read\_docx\_text, read\_resume

Reads a resume file and returns plain text.

**3.3 Resume–JD similarity**

python

CopyEdit

get\_tfidf\_similarity(text1, text2)

Calculates a **cosine similarity score** between resume and JD.

**3.4 Send prompt to LM Studio**

python

CopyEdit

call\_lmstudio(prompt, model, retries=3)

Sends a prompt to Zephyr model running on LM Studio (localhost:1234).  
Retries up to 3 times if the call fails.

**🧠 4. Main Pipeline: process()**

**✅ Step 1: Extract files**

python

CopyEdit

extract\_zip(raw\_resume\_zip, raw\_resume\_dir)

extract\_zip(jd\_zip, jd\_dir)

**✅ Step 2: Load resumes and JDs**

python

CopyEdit

raw\_resumes = { candidate\_name: resume\_text }

jd\_texts = { jd\_name: jd\_text }

Reads resumes and JDs into memory.

**✅ Step 3: Match resumes with best JD using TF-IDF**

python

CopyEdit

for candidate, resume\_text in raw\_resumes.items():

for jd\_name, jd\_text in jd\_texts.items():

score = get\_tfidf\_similarity(...)

keep best match

**✅ Step 4: Build prompt for LM Studio**

python

CopyEdit

prompt = f"""

You are a resume writing assistant...

"""

Includes raw resume and matched JD. Requests **summary, qualifications, experience, skills**.

**✅ Step 5: Call LM Studio**

python

CopyEdit

result = call\_lmstudio(prompt)

Gets response from Zephyr model.

**✅ Step 6: Extract info from LLM output**

python

CopyEdit

summary = re.search(...)

qualifications = ...

experience = ...

tech\_skills = ...

**✅ Step 7: Fill the EY template**

python

CopyEdit

final\_resume = ey\_template.replace("[SUMMARY]", summary)

Fills placeholders like [CANDIDATE\_NAME], [SUMMARY], etc.

**✅ Step 8: Save final output**

python

CopyEdit

with open(out\_path, "w") as f:

f.write(final\_resume)

Saves the final customized resume to a .txt file.

**▶️ 5. Run the code**

python

CopyEdit

if \_\_name\_\_ == "\_\_main\_\_":

process()

**✅ Summary of Workflow**

1. **Extract resumes and JDs**
2. **Read and preprocess text**
3. **Match each resume to best JD using TF-IDF**
4. **Send prompt to LM Studio with raw resume + JD**
5. **Get custom sections (summary, experience, etc.)**
6. **Insert sections into EY format template**
7. **Save as a new text file per candidate**

**BACKEND WORKFLOWS:**

**✅ Your Frontend Workflow (Stays the Same)**

1. **User provides a prompt** (e.g., “Create JD for X role”)
2. **User uploads their resume**
3. **User clicks "Customize Resume"**

**🔍 Now, for the backend, here are your 2 possible workflows:**

**✅ Option 1: Use ONLY Uploaded Resume + JD from Prompt (SIMPLE & FAST)**

* **Input:** Single resume + user prompt
* **Zephyr generates JD from prompt**
* **Zephyr customizes ONLY that uploaded resume using JD**
* **Does NOT refer to raw resume or JD folders**
* ✅ Recommended for one-click frontend integration

📌 **RAG is not needed** in this case, because you're not retrieving any previous documents — you're directly using the **prompt + uploaded resume**.

**🔁 Option 2: Use All Raw Resumes + JD Folder + RAG + Zephyr (Advanced)**

* **System searches across the Rawresumes/ folder** to find the most similar resume
* **Also matches best JD from JDs/ folder** using cosine similarity
* Optionally uses **RAG-style retrieval** from multiple JDs
* Then customizes based on the **best match**

❗ This is not required **unless** you're trying to batch-process multiple candidates **or** offer **“auto-match”** for recruiters.

**💡 So, for your specific frontend use-case:**

🔧 **You should go with Option 1** — simple Zephyr generation using uploaded resume + prompt-based JD generation.  
You do **not** need to refer to the rawresume/ or JDs/ folders unless:

* You're building a recruiter tool that compares among multiple candidates and JDs.
* You're doing batch customization.

**RAG-OLD**

Yes — integrating **RAG (Retrieval-Augmented Generation)** to use the **JDs and raw resumes stored in your backend folders** can significantly **boost the performance** and **intelligence** of your recruiter tool. Here's why:

**✅ What You Want:**

* **Frontend user provides:**
  + A **resume** (uploaded by user)
  + A **prompt** to generate a new JD (e.g., “Create JD for a Senior Data Scientist in healthcare with Python, GNN”)
* **Backend goal:**
  + Use **past resumes and JDs stored in the folders** to **inform and guide the LLM** (Zephyr) to:
    1. Generate **more realistic JDs** based on past examples
    2. Customise the uploaded resume to match the generated JD, using domain-specific language and success patterns.

**🚀 Why Using Stored Resumes + JDs (via RAG) Helps:**

**1. Richer JD Generation (via JD folder)**

* Your stored JDs act as **“examples of good JDs”**.
* When the user prompts Zephyr to generate a JD, the backend:
  + Retrieves top 3 similar JDs from the JD folder
  + Passes them as context to Zephyr’s prompt
  + Result: JD is **more grounded**, **domain-specific**, and **aligned to real expectations**.

**2. Smarter Resume Matching (via Rawresumes folder)**

* Instead of randomly picking the best match,
* Use your **past resumes** to:
  + Train or fine-tune matching heuristics
  + Improve Zephyr’s output tone and content with real resume examples
  + Enable **zero-shot or few-shot prompting** with similar candidates

**3. Improved Customisation Accuracy**

* When you ground Zephyr with past JDs/resumes:
  + It picks **more accurate keywords**
  + Adopts the **right voice, structure**, and **priority of skills**
  + Reduces hallucination and boosts **hiring-relevance**

**🧠 Analogy:**

Imagine trying to write a cover letter for a new job without seeing *any examples* vs having 10 cover letters for similar roles. That’s the difference RAG makes.

**⚙️ Implementation Suggestion (Backend):**

1. Use **FAISS** or **local semantic search** to index:
   * All .txt or .docx files in your JD folder
   * All resumes in rawresumes folder
2. When user provides prompt and resume:
   * Retrieve top 2–3 JDs and resumes **most similar** to the prompt and uploaded resume
   * Feed these into the Zephyr prompt (few-shot learning or prompt examples)

**🔁 Example RAG Prompt to Zephyr:**

sql

CopyEdit

Using the following examples of successful JDs and resumes, generate a new JD tailored to this user prompt:

User Prompt: "Create a JD for a ML engineer in credit risk domain using Python, XGBoost, SHAP"

Example JD 1: ...

Example JD 2: ...

Example Resume 1: ...

→ Generate new JD here.

**how the updated backend flow uses this frontend prompt:**

**“Create a JD for a retail IRB model development requirement. Mention that SQL is a must.”**

**🔄 Here's how the updated backend flow uses this frontend prompt:**

**✅ Step-by-Step Overview:**

1. **Frontend input:**
   * A user uploads a raw resume (e.g., AdityaVenkatesh.pdf)
   * They enter a prompt like: *"Create a JD for a retail IRB model development requirement. Mention that SQL is a must."*
2. **Backend flow (new version):**
   * Uses the prompt to:
     + Simulate JD context
     + Find similar JDs from JD folder (JDs/)
     + Find similar resumes from raw folder (Rawresumes/)
   * Sends everything to **Zephyr LLM** for response generation.
   * Replaces placeholders in the EY\_sample\_resume\_template.txt with:
     + JD-matched and prompt-aligned content
     + Raw resume name and sections
3. **Customization completes**, and a new file like AdityaVenkatesh\_customised\_resume.txt is saved in final\_customised\_resumes/.

**🧠 Why this works well?**

* **Prompt = synthesizes realistic JD context**
* **RAG = boosts Zephyr’s accuracy** by anchoring the model to your existing data
* **Frontend freedom + Backend grounding** = high-quality, role-specific resumes for recruiters

**💡 If you want to further enhance:**

* Store all prompts and resume names for **audit/history**
* Track **match scores** of each resume to each JD used
* Add **keywords detection** (e.g., SQL, IRB, PD, LGD, EAD, etc.) to improve semantic matching

Would you like me to wrap this entire logic into:

* ✅ A Flask API (/customise\_resume)
* ✅ A Streamlit or React UI for easy testing?

**ACCURACY CALCULATION:**

**✅ 1. Exact JD Created by the Code (via Zephyr):**

The code uses this **frontend prompt**:

*“Create a JD for a retail Scorecards model development requirement. Mention that SQL is a must.”*

From the customized resume output in Nidhika-Tomar\_customised\_resumenidhika1.txt, we can infer that **Zephyr synthesized** the following **JD-style information** (even though the JD text itself is not explicitly printed, it is reflected in how the resume was customized):

* **Role Focus**: Retail Scorecards model development
* **Technical Emphasis**: SQL is mandatory
* **Domain**: Risk management, secured portfolios, capital emulators, RWA requirements
* **Soft Skills and Experience Requirements**: Team collaboration, experience with regulatory and statistical modeling

This JD is embedded and **implicitly applied** in how the SUMMARY, EXPERIENCE, and TECHNICAL SKILLS sections are rewritten.

**✅ 2. How to Measure Accuracy of Customization:**

To **measure the accuracy of resume customization** based on the synthesized JD, use these **4 criteria**:

**📌 A. Relevance Match (Qualitative Manual Check):**

Check whether:

* Skills from the JD (e.g., SQL, scorecards, retail model development) are reflected in the customized resume.
* Summary and Experience sections are rewritten in JD-aligned language.

✅ *Example from Nidhika’s resume*:

*“Her expertise in the modeling of secured portfolios, capital emulators, impairment estimators, and RWA requirements…”*  
➡ Matches the JD’s implied expectation.

**📌 B. TF-IDF or BERT Embedding Cosine Similarity:**

You can **numerically evaluate alignment** by calculating cosine similarity between:

* Synthesized JD
* Customized Resume

This uses TfidfVectorizer or SentenceTransformer (BERT embeddings). Scores closer to **1.0** show higher alignment.

**📌 C. Section-wise Match Audit:**

Evaluate each section of the resume:

* **SUMMARY**: Tailored summary aligned to JD theme?
* **QUALIFICATIONS**: Educational details preserved?
* **EXPERIENCE**: Rewritten to reflect JD keywords (e.g., retail scorecards, stress-testing)?
* **TECHNICAL SKILLS**: SQL explicitly mentioned?

✅ *From the resume*: SQL is highlighted, and scorecards & secured book models are detailed.

**📌 D. Named Entity and Keyword Coverage:**

Use NLP tools like spaCy to:

* Extract entities and keywords from the JD.
* Count how many of these appear in the customized resume.

This gives a **keyword coverage ratio**:

ini

CopyEdit

coverage\_score = (Number of JD keywords in Resume) / (Total JD keywords)

**RAG:**

**RAG or historical data is *not required*** for resume–JD customization **to work**, but using it can **significantly improve** the quality and accuracy. Here's the honest breakdown:

**✅ When RAG or historical data is not required:**

The current approach (your code) is:

* Simple, fast, and works well **if**:
  + The resume is already rich and relevant
  + The JD prompt is clear and specific
  + The LLM (like Zephyr-7B) has enough general knowledge to match concepts

**Your code is enough** to:

* Generate a basic JD from a prompt
* Align a resume with that JD using only what's in the resume
* Output structured, template-aligned content

**🔍 When RAG or historical data becomes helpful or recommended:**

1. **For context-rich customization**:
   * Example: If the resume is **sparse**, RAG can pull matching project details or past resumes to enrich the output.
2. **For consistent tone and formatting**:
   * Example: Pull from historical successful resumes at EY or by the same candidate for reuse.
3. **For domain-specific job understanding**:
   * Example: In finance, "PD/LGD modeling" might mean different things across banks. A RAG system could reference internal documentation.
4. **For real-time retrieval of latest JD formats**:
   * Instead of synthetically generating a JD, RAG could pull **actual current JD data** from job portals or a stored library of past JDs.

**🧠 Without RAG — what’s the limitation?**

* Your code relies **entirely on the LLM's training knowledge + resume content**
* If the model lacks understanding or the resume is weak, the customization will be shallow or inaccurate

**🟢 When to stick with current approach:**

* You’re doing fast prototyping or automation at scale
* You don’t have access to vector databases or past resume-JD pairs
* You trust your resume is complete and accurate

**🟡 When to consider adding RAG:**

* You have a database of real JDs or resumes
* You want to retrieve project details or role descriptions dynamically
* You want to personalize across multiple versions of the same resume for different JD prompts

**ChatGPT said:**

Yes, the code I provided **already includes logic for all four accuracy steps**, though the **manual part (A)** needs to be performed through visual inspection by the reviewer. Below is a clear mapping between your 4-point checklist and how each is addressed in the code:

**✅ Final Confirmation: Does the Code Cover All 4 Steps?**

| **✅ Step** | **Description** | **Covered in Code?** | **Details** |
| --- | --- | --- | --- |
| 📌 **A. Relevance Match (Qualitative Manual Check)** | Human must read and verify if resume summary & experience align with JD | ✅ **Manually Printed** for Review | The code **prints the generated JD and final resume**, so you can manually verify if SQL, scorecards, retail model phrases are reflected. |
| 📌 **B. TF-IDF or BERT Embedding Similarity** | Cosine similarity score using TF-IDF | ✅ Yes | Implemented via compute\_tfidf\_similarity() using TfidfVectorizer. |
| 📌 **C. Section-wise Match Audit** | Check presence and quality of sections like SUMMARY, EXPERIENCE, etc. | ✅ Yes | section\_audit() returns section flags like ['summary', 'qualification', 'experience']. |
| 📌 **D. Named Entity / Keyword Coverage** | Ratio of JD keywords appearing in resume using SpaCy | ✅ Yes | keyword\_coverage() extracts keywords using spacy and computes ratio + matched words list. |

**✏️ Suggestions for Enhancement (Optional)**

* **A: Manual relevance** is necessarily subjective and must be **done by you** after seeing the printed generated\_jd and final\_resume.
* **B: Optionally**, you can replace TfidfVectorizer with a **BERT model** like sentence-transformers/all-MiniLM-L6-v2 if you're able to run sentence-transformers.
* **C: Section audit** could be improved to verify **specific keywords inside sections** (e.g., check if "SQL" appears only inside TECHNICAL SKILLS).
* **D: You could print top unmatched keywords** from JD that didn't appear in the final resume if you'd like.

**FIXES-EMPTY-NIDHIKA**

**Steps to Fix and Make the Code Robust**

**✅ 1. Force the LLM to output consistent section headers**

Modify your prompt to **enforce exact headers**, like this:

text

CopyEdit

Use EXACTLY these section headers, in this order, all in UPPERCASE:

SUMMARY:

QUALIFICATIONS:

EXPERIENCE:

TECHNICAL\_SKILLS:

**✅ 2. Add a fallback print to debug the raw LLM output**

Before regex parsing, log the LLM output:

python

CopyEdit

print("===== LLM OUTPUT START =====")

print(result)

print("===== LLM OUTPUT END =====")

**✅ 3. Make regex case-insensitive and flexible**

Here’s an improved extractor using better patterns:

python

CopyEdit

def extract\_section(text, section\_name, next\_section=None):

try:

pattern = rf"{section\_name}[:\s]\*\n\*(.\*?)(?=\n\*{next\_section}[:\s]\*\n\*|$)" if next\_section else rf"{section\_name}[:\s]\*\n\*(.\*)"

match = re.search(pattern, text, re.IGNORECASE | re.DOTALL)

return match.group(1).strip() if match else ""

except Exception as e:

print(f"Error extracting section {section\_name}: {e}")

return ""

Use:

python

CopyEdit

summary = extract\_section(result, "SUMMARY", "QUALIFICATIONS")

qualifications = extract\_section(result, "QUALIFICATIONS", "EXPERIENCE")

experience = extract\_section(result, "EXPERIENCE", "TECHNICAL\_SKILLS")

tech\_skills = extract\_section(result, "TECHNICAL\_SKILLS")

**✅ 4. Use print checkpoints before saving**

Ensure all values are non-empty before writing the file:

python

CopyEdit

print("Summary:\n", summary)

print("Qualifications:\n", qualifications)

print("Experience:\n", experience)

print("Tech Skills:\n", tech\_skills)

**✅ Optional: Validate with Keyword Presence**

Use a quick check before saving:

python

CopyEdit

if not any([summary, qualifications, experience, tech\_skills]):

print("⚠️ All extracted fields are empty. Check model output and prompt formatting.")

return

**✅ When RAG or Historical Data Is Not Required**

The current **Zephyr-based code works well without RAG** if:

* The resume is already **rich and relevant**
* The **JD prompt** is clear and specific
* The **LLM (like Zephyr-7B)** can understand the domain context

**Your code is capable of:**

* Generating a JD from prompt
* Customizing a resume based only on input resume and JD
* Producing clean, EY-template-aligned resumes

**🔍 When RAG or Historical Data Is Helpful or Recommended**

1. **Context-Rich Enrichment**
   * Resume lacks details → RAG can pull similar project experiences, prior versions
2. **Tone and Formatting Consistency**
   * Use historical resumes from EY or same candidate to standardize structure
3. **Domain-Specific Understanding**
   * RAG can differentiate “PD/LGD modeling” usage across banks or roles
4. **Real-Time Job Data Retrieval**
   * Pull actual JDs from job portals or internal libraries instead of synthesizing

**🧠 Limitations Without RAG**

* Entirely relies on:
  + Resume quality
  + Prompt clarity
  + LLM’s pre-trained knowledge
* May lead to **generic or shallow customization** if inputs are weak

**🟢 Stick With Current Approach If:**

* Fast prototyping or automation at scale
* No access to past resumes/JDs or vector DBs
* Resume is already strong and LLM-savvy

**🟡 Consider Adding RAG If:**

* You have a JD/resume repository
* Need to dynamically fetch domain-specific content
* Want to personalize for multiple target JDs
* Yes — here's a **clear breakdown** answering your questions:
* **✅ Q1: Is the line to print JD included?**
* Yes — **line 97** (approximately) prints the JD that was retrieved using RAG:
* python
* CopyEdit
* print("✅ JD Retrieved From:", jd\_path)
* And the actual JD content is used in this line within the prompt:
* python
* CopyEdit
* Job Description:
* {enriched\_jd}
* So while it’s not printed verbatim to avoid clutter, you can add this right after the retrieval if you want to **print the full JD content**:
* python
* CopyEdit
* print("\n📄 Retrieved JD Content:\n", enriched\_jd)
* **✅ Q2: Which lines of code implement RAG?**
* These are the exact **RAG-related lines**:
* **🟩 1. Retrieving the JD from your local JD repository:**
* python
* CopyEdit
* jd\_path, enriched\_jd = get\_best\_match(prompt\_text, jd\_folder)
* This uses cosine similarity (TF-IDF) to retrieve the most relevant JD instead of generating it synthetically.
* **🟩 2. If resume is sparse, augment it with past resume data:**
* python
* CopyEdit
* if len(resume\_text.strip().split()) < 150:
* \_, enriched\_resume = get\_best\_match(prompt\_text, resume\_folder)
* resume\_text += "\n\n[ENRICHED CONTEXT]\n" + enriched\_resume
* This pulls from historical resumes in your Rawresumes folder when the current resume lacks enough content.
* **🟩 3. The get\_best\_match() function (core RAG logic):**
* python
* CopyEdit
* def get\_best\_match(input\_text, folder, top\_k=1):
* ...
* tfidf = TfidfVectorizer().fit\_transform([input\_text] + contents)
* sims = cosine\_similarity(tfidf[0:1], tfidf[1:]).flatten()
* best\_idx = sims.argmax()
* return paths[best\_idx], contents[best\_idx]
* This performs local **retrieval from your JD/resume repositories** using TF-IDF and cosine similarity — classic RAG.
* **✅ Summary**

| **Feature** | **Code Line(s)** |
| --- | --- |
| ✅ JD Retrieval via RAG | jd\_path, enriched\_jd = get\_best\_match(...) |
| ✅ Print JD path | print("✅ JD Retrieved From:", jd\_path) |
| ✅ Optional: Print JD text | print("\n📄 Retrieved JD Content:\n", enriched\_jd) |
| ✅ Resume Enrichment via RAG | if len(resume\_text.strip().split()) < 150: |
| ✅ Core RAG Logic | Entire get\_best\_match() function |

**🏆 Best, Easy & Quick Method:**

**✅ Streamlit Standalone App + Deploy on Streamlit Cloud (Free)**

**✅ Why Streamlit?**

| **Feature** | **Benefit** |
| --- | --- |
| 🧠 Built for ML/AI apps | Perfect for file upload + text input + displaying results |
| ⚡ Super fast to build | You can get a working app in **under 1 hour** |
| 🌐 Free Hosting | Streamlit Community Cloud allows free public deployment |
| 🔒 Internal Usage | You can protect it using Google Auth or basic streamlit login hacks if needed |

**✅ What It Will Do**

* st.text\_area() → Users type the JD prompt
* st.file\_uploader() → Users upload their resume
* Code behind will:
  + Call call\_lmstudio()
  + Run customise\_resume(prompt, file)
  + Show output + metrics + download button

**✅ Deployment Steps**

1. Push the code to **GitHub**
2. Go to https://streamlit.io/cloud
3. Connect your GitHub → Select repo → Deploy 🚀

That’s it! You’ll get a sharable public link.

**✅ What You Get**

* ✅ Web app with prompt + file upload
* ✅ No Flask or HTML required
* ✅ One file (app.py) = UI + Logic
* ✅ Easy sharing across your team
* ✅ Free forever (as long as public app is OK)

**⚠️ One Note:**

Your current code uses LM Studio running **locally** (localhost:1234).  
To use it from Streamlit Cloud, either:

* Expose LM Studio via public IP (⚠️ not ideal), or
* **Replace LM Studio with OpenAI/GPT API** (if you have an API key)

**🔧 Want Me To Build It?**

I can give you a ready-to-run:

* ✅ app.py (Streamlit code)
* ✅ requirements.txt
* ✅ GitHub-ready setup for Streamlit Cloud

**Would you like the Streamlit code using:**

* LM Studio (for local testing only)  
  or
* OpenAI API (if you want to host it online)?