Annexure-B

Corrigendum to Pre Bid Queries and Replies for to RFP No. CCTNS (103)/CDA-R/45/NCRB/2019 dated 13/05/2020 for selection of Agency for Development & Implementation of ML & NLP Based Analytics on FIR data

S. No	Clause No	Clause Point No.	RFP Page No.	Original Clause	Clarification Sought	NCRB Response
1. Trident information Systems Pvt Ltd						
1	4	4.6	6	Bidder should use open source technologies for development of the system.	Please consider all available technology.	All Open Source Technology will be considered which means NCRB does not have to pay any license fee or annual support charges.
2	4	4.4	6	NCRB will provide the keywords-based decision matrix to tag the existing FIR records to different categories/sub-categories/sub-sub-categories which will be used for training of NLP based ML System.	Decision Matrix – what is this? Can we see a sample?	Based on certain keywords that are there in the text of FIR, the decision matrix will indicate what type of crime methods has been used. Such a decision matrix would be evolved together with the selected vendor.
3	4	4.10	6	Bidder has to provide the simple GUI to search the results of classification based on User Inputs	What is GUI?	GUI is the Graphics User Interface on which results will be displayed.
4	5	5.4	7	Bidder shall have minimum two resources who have experience of Police System and ML/AI/NLP.	Please consider experience in paramilitary forces also.	Experience of Paramilitary will be considered.
5	8	8.1	8	No mention of advance/mobilisation amount.	NCRB should consider some advance/mobilisation amount to bidder.	No advance/mobilisation amount to bidder will be considered.

6	11	11.1, 11.2 & 11.3	9	1. Any delay attributed to bidder shall attract a penalty of 1% of contract value per week will be applied. 2. Maximum penalty under this contract will be capped at 10% of contract value. 3. Payments links with accuracy of FIR categorization. i. 90% and above - No penalty ii. 80% and above but below 90% - 5% iii.70% and above but below 80% - 10% iv. Less than 70% - No payment and contract terminate	Since this is new technology, penalties should be reviewed. We recommend followings: i. 70% and above accuracy - No penalties ii. 60% to 70% accuracy - 2% iii. 50 to 60% accuracy - 5% iv. Max 5% penalties instead of 10% and termination at less than 50% accuracy.	As per RFP; No change in penalties will be considered.
7	11	11(B)	10	Tender Document	The accuracy of categorization of FIRs will be checked & validated as per below formula: The 50k initial set of documents for each language i) is this pre-labelled data? Ii) Will this be given to us for training?	I) Yes, NCRB will check the accuracy based on its pre-labelled classification II) NO, as this data will be used to check the accuracy of supplied application.
8	12	12(b)	11	EMD 2.5 lakh in the form of DD/Pay Order in the name of "AO, PAO, DCPW" payable at Delhi.	EMD should be exempted for MSMEs	As per Govt. Norms; Please enclose relevant documents for exemption.
9	How to query CCTNS system? Is there an API?					Direct query or API can be used with read only access to access CCTNS system. API will be developed by the selected solution provider
10	Is the free flowing FIR data available as text or some other format?					Text format in digital format
11	Like Hindi, will there be CDAC APIs for the other 7 languages?					Yes

12	Is it transliteration (खून to "Khoon") or translation (खून to "Murder")	Transliteration is (खून to "Khoon")
13	Watson Discovery is an Industry leader offering for NLP & NLU and built by IBM using open source technologies	No Comments

2. Amvar Data Tech Pvt Ltd

14					1. Is this complete final tender for purchase or is it only a selection procedure ? In other words there a already budgeted mandate for Purchase or is this to create a budgetary estimate. It mentions the EMD amount of INR 250,000/- so it seems to be a tender submission. Hence the above question.	This is a Tender for Selection of Agency for Development & Implementation of ML & NLP Based Analytics on FIR data on payment basis
15					2. We are an established Analytics company working over the last four years and have executed several projects in Big Data/ Machine Learning/ Computer Vision and NLP through our marketing partner in the USA. Can we quote these projects as evaluation criteria or we need to show the work orders from our US counterparts?	Work orders/Documents establishing the bidder's credentials having executed similar projects have to be provided.
16					3. Will projects completed with Private sector in US or India can demonstrated for achieving scores as given in Section 6?	Documents in the bids can be enclosed, the evaluation committee will verify and if found suitable will be considered.

17					4. Will adequate sample data from CCTNS system be provided to us for Demonstration as proposed in Sr. No 3 of Section 6?	Bidders will be provided sample data after anonymization to protect the privacy by removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous
18					5. Can an Indian entity to a joint bidding with an US company for this project. Will it be considered a valid bid?	Pre-qualification criteria don't explicitly debar an Indian entity doing joint bidding with an US company.
19					6. Can we attend the Pre bid Conference on 3rd June remotely over phone/ video in view of the current situation in most cities of the country?	Yes
3. Cartesian consulting						
20					1. Data Extraction: How to query CCTNS system? Is there an API?	As per serial no. 9
21					2. Data format: Is the free flowing FIR data available as text or some other format? Can we get a sample to look at?	Data is in Text format; Bidders will be provided sample data after anonymization to protect the privacy by removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous
22					3. Data quality: Are there any data quality issues in the source data that we already know of? *(Most NLP solutions live and die by the quality of their pre-processing pipeline. The worse*	Yes, there may be some data quality issues which have to be addressed by the selected vendor in its NLP model.

						the data quality, the more the pre-processing work required.)	
23						4. Translation vs transliteration: Is it transliteration (खून to "Khoon") or translation (खूनto "Murder"). The reference to the CDAC library uses both words in different places.	Transliteration is (खून to "Khoon"). However "खून" may mean Murder or Blood depending on the context in FIR.
24						5. Decision Matrix: What is this? Can we see a sample?	As per serial no. 2
25						6. Test data: The 50k initial set of documents for each language •Is this pre-labelled data? •Will this be given to us for model training?	As per serial no. 7
26						7. Static problem or not: Are there subcategories/categories that we have not been given (or aren't present in the data but might occur in the future), do we work with the assumption that the model has to be flexible enough to accommodate new classes over time? Or do we think of this as a static problem?	Whenever a new Crime Head/Modus Operandi (MO) is added, the system should be able to add keywords to learn and train itself for the new crime head/MO

27					8. Long term maintenance: We have seen that in most cases a solution like this requires maintenance. Once the system is built and deployed, we are assuming that it requires someone to maintain it and update/improve the models over time. Is it ok to assume model maintenance is part of the scope?	No, model maintenance is not in the scope of the RFP but the selected vendor has to provide 5 Years warranty of the solution.
28					9. Other languages: This bid is for English and Hindi only and other 7 languages model builds will be a scope extension later (after the English and Hindi are successfully deployed). Is this understanding correct? Also like Hindi, will there be CDAC APIs for the other 7 languages?	As per RFP current scope of tender is only for Hindi and English FIRs. Other languages will be considered in the future.
4. SAS Institute (I) Pvt Ltd						
29		4.1	6		Please confirm if 1 Lac records of each language is the overall number during 5 years? If it is for 5 years then what is the present data size and if it is the current data size what is the expected year on year increment?	Vendor has to work only for 1 lac records of each language (Hindi and English) which will be provided by NCRB from its complete database.

30		4.3	6		Please confirm if the CDAC will be supplying transliteration API or translation API. The difference between the two is that transliteration will just change the word to English ("अजय यहां है" will be converted to "Ajay yaha hai") while the translation will be able to convert hindi readable text to English ("अजय यहां है" will be converted to "Ajay is here").	CDAC transliteration engine is provided free of cost to vendor for usage. Bidders can check accuracy of the engine. If Bidder is using any other language transliteration engine then that should be provided free of cost to NCRB forever. CDAC Transliteration Engine will result in ("अजय यहां है" will be converted to "Ajay yaha hai")
31		4.3	6		Since SLA penalties are linked to accuracy, it is necessary to understand accuracy of CDAC provided APIs. It is requested to provide some estimates on accuracy.	As earlier, Serial No. 30; Vendor can visit NCRB and assess the accuracy.
32		4.6	6		We request that the bidder should be free to use open source/COTS based solution to achieve the desired outcome.	Please refer serial no. 1
33		4.10	6		Does the department require following search capabilities: Boolean Search Fuzzy Search Phonetic Search Synonym Search Wildcard Search Query Builder Form Search Proximity Search	Vendor is free to use the desired algorithm/logics to provide the expected results as per RFP.

34					It is assumed that for other languages also CDAC translation APIs will be provided in the future.	Yes, but the current scope is only for English and Hindi
35					We believe that the NLP capability should not be limited to categorization and classification but should also support entity detection, relationship extraction, topic extraction and coreference resolution.	Current scope is limited to categorization of FIRs
36					NLP capabilities should include automated parsing, tokenization, part-of-speech tagging, and lemmatization. These would help in information extraction from raw text.	As earlier, Serial No. 33
37					Will the department require capability to add predefined concepts to extract common entities, such as names, organizations, locations, expressions of time, dates, quantities, percentages and more	As earlier, Serial No. 33
38					Should misspelled analysis be part of the solution to be able to detect misspelled words and offer set of variants that include the properly spelled word?	As earlier, Serial No. 33

39					Will the department require automatic topic discovery capability through unsupervised machine learning methods such as Latent Drichlet algorithm?	As earlier, Serial No. 33
40					Will the department require capability to merge or split system generated topics to create user-defined topics?	As earlier, Serial No. 33
41					We believe that the solution should also support rule generation for document categorization. This will improve the efficacy of the categorization where not much data is present for AI ML training.	As earlier, Serial No. 33
42					Does the department like to have the ability to visualize extracted entities, facts and relationships using network diagrams?	It is not in the scope of the RFP
43					Will the department like to extract entities such as victim, convicts etc. from the FIR data?	do
44					Is it required to perform entity resolution to create golden records of individuals? Ex. RK Srivastav arrested for burglary in FIR no x is same as Ram Kumar Shreevastava arrested in FIR no y. So golden record with	do

						Ram Kumar Srivastav will be created linking both the FIRs.	
45						Once a golden record is created for individual, does the department wants to enrich the record with individual details like photo, physical appearance, family and gang association, mobile details, external data (like VAHAN, passport etc.)	do
46						To visualize golden records, does the department require a single integrated screen to view entities and its attributes along with all the linked records? This would create dossier of individuals.	do
47						Will it be beneficial for the department to have a network analytics capability to explore relationship amongst different entities?	do
48						Does the department require alerting capability? To be able to configure alerts through a GUI with scoring models and trigger conditions.	do
49						Will it be required to export data through raw data dump or templatised reports?	NCRB will provide the uncategorised FIR data in the form of text or through read only access to the database. After the categorization, the selected vendor will upload the data at the desired location/in the desired

							database as per NCRB. Details will be discussed with the selected vendor.
5. C-DAC Noida							
50		Section 12(a) Page 10				CDAC is the Central Govt. Organization / Society, exemptions for EMD may be allowed	As Serial No.8
51		Section 4 Page 5				The system will be developed on the central database. FIR database with the ability to comprehend free-flow text data such as field 12 of CCTNS' FIR data of CCTNS **Expectation:** Please name the database and confirm that field 12 details are stored in database itself.	Yes. MySQL/MSSQL/Sybase Field 12 details are stored in the database
52		Section 5 Page 6				Bidder shall have minimum two resources who have experience of Police System and ML/AI/NLP. Bidder to provide the short resume of resources. Customer may meet the people or check the project references **Expectation:** It is understand that bidder can provide two	Yes

					SAMPLE INDICATIVE RESUMEs to verify the organization capability in the area. But, at the project execution, same persons or similar/ better person can be deployed. Kindly confirm.	
53		Section 3 (10) Page6			It says 90 calendar days (i.e. ~ 3 months). It is assumed that this period is for the two languages in scope. How about other languages and other scope of the project, will it be considered now by bidders or will there be a separate tender. Kindly confirm above with the overall total time period of the project.	In current RFP, current scope is two languages only.

6. NEURIOT TECHNOLOGIES LLP

54				12. Additional Documents to be submitted with bid response	a) Bid Signing Authority Letter. **Expectation:** "Considering the COVID-19 situation in the country and restricted movement of people and touching of items, could you please consider giving the option of paying EMD through RTGS bank transfer? We will absorb any bank charges applied to us for this transfer. We will attach the proof of	Physical/Scanned copy of Bank Guarantee can be given with bid and the physical copy can be given during evaluation

					payment in the tender response document. Request you to please provide the bank details."	
55					b) EMD 2.5 lakh in the form of DD/Pay Order in the name of "AO, PAO, DCPW" payable at Delhi."	Yes
7. IBM India Pvt Ltd						
56		4 - Scope of Project		4.6 - Bidder should use open source technologies for development of the system	Can we propose products offering that has been built using open source technologies and that can be hosted on-Prim	As per Serial No. 1
57		4 - Scope of Project		General	Is there any weightage on the product being proposed that is part of any industry analyst report like Gartner, forester etc.	No
58		5 - Pre-Qualification Criteria		4. Bidder shall have minimum two resources who have experience of Police System and ML/AI/NLP.	Can OEM resources be accepted along with the bidder	Resources should be from the company which is bidding
59		5 - Pre-Qualification Criteria		2. Bidder should have executed 5 projects in the area of Machine Learning or Artificial Intelligence or Natural Language Processing	Can OEM experience be considered along with bidder	Experience of the bidding company will be considered
60		4 - Scope of Project		General	How to query CCTNS system? Is there an API?	As per Serial No. 9

61		4 - Scope of Project		General	Is the free flowing FIR data available as text or some other format?	Digital text
62		4 - Scope of Project		4.3 For language translation (from local languages to English), NCRB will provide the CDAC supplied APIs for transliteration of text.	Like Hindi, will there be CDAC APIs for the other 7 languages?	Yes, but the current scope is 2 languages Hindi & English only.
63		4 - Scope of Project		General	Is it transliteration (*खून to "Khoon"*) or translation (*खून to "Murder"*)	As per Serial No. 12
64		4 - Scope of Project		General	Decision Matrix –what is this? Can we see a sample?	As per Serial No. 2
65		4 - Scope of Project		General	The 50k initial set of documents for each language •is this pre-labelled data? •Will this be given to us for training?	As per Serial No. 7
8. Roscoe Lobo (Leader-Data & AI)						
66				CCTNC	What is the back-end storage / architecture for CCTNS?	Data is stored in MySQL/MSSQL/Sybase and CCTNS has SOA architecture
67				CCTNC	Is the data to be queried every time information / records are expected to be fetched from CCTNS? Or are they to be stored as part of the solution components where processing will take place?	Details will be finalised with the selected vendor.

68					CCTNS	How to query CCTNS system? Is there an API?	As per Serial No. 9
69					CDAC	Is there flexibility to use non-CDAC APIs for Language Translation?	As per Serial No. 30
70					CDAC	Are CDAC APIs going to be available for "transliteration" or "translation"?	Transliteration
71					CDAC	Like Hindi, will there be CDAC APIs for the other 7 languages?	Yes, but it is not in the current scope of work
72					CDAC	Will CDAC APIs facilitate "transliteration" (i.e.: खून to "Khoon") or "translation" (i.e.: खून to "Murder")?	As per serial No. 12
73					DB	Which Database is going to store the categorized/classified information? Do we need to consider proposal for database licenses and sizing from our side?	No, if required by your solution please use an open source database.
74					Decision Matrix	What do you mean by keyword-based decision matrix? Can you share a few examples?	As per Serial No. 2
75					Decision Matrix	How many levels of depth/hierarchy will the keyword-based decision matrix have?	It is dynamic, depending up on the crime.
76					Decision Matrix	Decision Matrix – what is this? Can we see a sample?	As per Serial No. 2

77					Records	Which years do the 1L (total for Hindi & English) records pertain to? How many more year are to be considered for a futuristic and scalable solution sizing?	NCRB has FIR data of more than 10 years and data records are huge in number. However, for this tender, we will provide 1 Lac records each of Hindi and English languages which could be from multiple years.
78					Records	Is the free flowing FIR data available as text? or some other format?	Yes in digital text.
79					Records	The 100k initial set of documents for each language - will this data be pre-labelled?	As per serial no. 7
80					Records	The 100k initial set of documents for each language - will this be given to us for training?	Yes to the selected vendor only in English and Hindi
81					Technology	Watson Discovery is an industry leader offering for NLP & NLU and built by IBM using open source technologies & can also be leveraged by multiple other open source technologies such as R, Python etc. using Watson APIs. Hope this would qualify under technology requirement?	As per serial no. 1. Qualification of any solution will be decided by the evaluation committee

9. PricewaterhouseCoopers Private Limited						
82	4	4.9	6		Current Clause states that Bidder will be provided necessary H/W with required software/ database and storage where bidder will be able to store the NLP/ ML based analytics data. Database or any other S/W tools will be part of the bid. Need clarity on the clause. Will the bidder be provided necessary H/W as per their requirement or as desired by NCRB. Need clarity. As the clause mentions, the database or any other S/W tools will be part of the bid, the minimum specifications of the tools need to be defined/specified in the bid, as the cost may vary.	For acceptance test/deployment of the solution, NCRB will provide the deployment infrastructure (servers, storage, Operating System, virus scanner, network, C-DAC supplied Language Transliteration API). However if the supplied solution is using any licensed software/item then it will have to be provided as part of bid cost for the unlimited use period by NCRB.

83	4	4.11	6		NCRB will use its existing resources to development of reports/graphs on the processed data from this system. Please provide the details of the tools available with NCRB for data visualisation and would developing reports on those tools be part of the bidder's responsibilities	Current scope is categorization of FIRs only.
84	5	5.2	7		The pre-qualification requirement of minimum **5** projects in ML, AL, or NLP is leading to maximum score in Technical The minimum no of projects in PQ should be limited to 2 projects for qualifying.	No Change

85	5	5.2	7		Current Clause states that Bidder should have executed 5 projects in the area of Machine Learning or Artificial Proposed clause. Bidder should have executed or undertaking 2 projects in the area of Machine Learning or Artificial Intelligence or Natural Language Processing Intelligence or Natural Language Processing	No Change
86	5	5.2	7		No Eligibility criteria of minimum turnover defined	As per the RFP
87	5	5.4	7		When would the POC demonstration take place and who shall provide the data and environment for the demo? Need Clarity on the same.	Demo/POC will be done by bidder during technical evaluation phase and NCRB will inform bidders in advance. However, for demo/POC all items need to be brought by bidder at their cost/risk. Bidders will be provided sample data after anonymization to protect the privacy by removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous

88	6	6.2	7		Current Clause states: Number of ML/NLP/AI Projects executed: 4 Points for each project subject to maximum 20 points The pre-qualification requirement of minimum 5 projects in ML, AL, or NLP is leading to maximum score in Technical (20 marks).	No Change
89	6	6.3	7		Current Clause states that Presentation/ Demo of the Proposed System: - Presentation of Proposed System to be made in person by the bidder (Maximum 20 marks). - Demo of the proposed technology applied on Crime Data (Maximum 10 marks).	No Change

| 90 | 6 | 6.4 | 7 | | Current Clause states:
a) Minimum 5 Years of Experience (Maximum 5 marks)
b) Technology experience of Machine Learning, Natural Language Processing and Artificial Intelligence (Maximum 10 marks)
c) Domain experience of applying relevant technologies in Crime Analytics (Maximum 10 marks)

Need clarity. Is this clause to be read in conjunction?
If yes, is the requirement of minimum 5 years in Tech profile or domain profile separately or 5 years combined i.e. tech profile in desired domain?
If not, then the experience requirement is not clearly defined, as below:
a) Are we talking about total experience of resource, that may or not be in ML/AL/NLP or in desired domain".
b) Technology experience in no of years is not defined.
c) Domain experience in no of | 2 Proposed Resources if have minimum 5 years experience will get 5 Marks

2 Proposed Resources if have Technology experience of Machine Learning, Natural Language Processing and Artificial Intelligence will get 10 Marks

2 Proposed Resources if have Domain experience of applying relevant technologies in Crime Analytics will get 10 Marks

Domain Experience is to be read as Police Domain Experience |

					years is not defined. Also, the experience category defined in PQ is Police System and in TQ it is domain exp in Crime Analytics. These categories are not same. Need clarity	

| 91 | 6 | 6.4 | 7 | | Current Clause states that "Minimum 5 Years of Experience (Maximum 5 marks)"

Qualifying experience requirement is fetching maximum marks in experience. Suggested clause:
"
Exp less 2 years of experience: 0 Marks
Exp between 2-3 years of experience: 2 Marks
Exp between 3-4 years of experience: 3 Marks
Exp between 4-5 years of experience: 4 Marks
Exp more than 5 years of experience: 5 Marks
" | No Change |
| 92 | 6 | 6.4 | 7 | | Total Maximum Marks after summing up all sections is coming out to be 30+30+30+30 = 110 but the total Max marks mentioned is 100.

Need to be corrected. | Maximum Technical score is 110 instead of 100 as mentioned at section - 6 of RFP. Technical evaluation will be done out of 110 marks. Later during calculation the mark obtained out of 110 by a bidder will be normalized on the scale of 100 and will be used in section 9 (b) of RFP.

Additionally, a vendor has to obtain 60% as the technical |

						score for qualifying for the opening of its commercial bid.
93	6	6.4	7		Domain experience of applying relevant technologies in Crime Analytics (Maximum 10 marks) Similar to point 6, separate resources with expertise in analytics and crime should be considered	No Change
94	8	8.1	8		Development and Deployment of System -On delivery of software, installation of the same on NCRB server Would the bidder also be developing the NLP system or only supplying the software, please clarify	Scope is the development and implementation of ML and NLP based analytics on FIR data.
95	9	9.b	8		Formula for evaluation score = 60 * (points of bidder/total technical points (i.e. 100) + 40*(price of L1 bidder/price of bidder). Request to update the QCBS weightage to 80:20 from current 60:40 of technical to financial score	1. Maximum Technical score is 110 instead of 100 as mentioned at section - 6 of RFP. Technical evaluation will be done out of 110 marks. Later during calculation the mark obtained out of 110 by a bidder will be normalized on the scale of 100 and will be used in section 9 (b) of RFP. Additionally, a vendor has to obtain 60% as the technical

						score for qualifying for the opening of its commercial bid. QCBS evaluation criteria as per RFP. 2. No change in QCBS weightage
96	11	11.2	9		Accuracy in categorization FIRs both in Hindi and English of 90% and above With the capabilities of currently technology achieving high accuracy is difficult, a relaxation should be provided in accuracy of 90% and the limit reduced to 75%	No Change

10. Inspira Enterprise India Pvt. Ltd.

97	2		5	The required solution must automatically classify the FIRs under different Crime category/sub-category/sub sub-category based on the information stored in free-flow text of FIRs with the help of Machine Learning and Natural Language Processing (NLP) methods & technologies	Request NCRB to please elaborate on the nature of free flow text is it going to Handwritten/Digital. It would be great if you can share one sample each of English and Hindi FIR masking critical information	Free flow text is digitally stored in relational database. NLP has to be applied to a column containing FIR text which contains how, where, when and why a crime has occurred and who all are involved in the crime. Bidders will be provided sample data after anonymization to protect the privacy by removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous

98	3		5	Date of pre-bid meeting	Request NCRB, if Pre bid Meeting could be virtually scheduled as currently due to Covid situation travel would be a challenge for concerned bidders to attend the prebid meeting.	considered
99	11		9	SLA & Penalties	Request NCRB to please consider below Penalty 85% and above - No penalty 75% and above but below 85% - 5% 65% and above but below 75% - 10% Less than 65% - No payment and contract terminate	As per serial no. 6
100				General Query	Please clarify response time, resolution time to be mentioned in SLA in detail for the warranty period.	Will be decided mutually with the selected vendor
11. KPMG Advisory Services Private Limited						

| 101 | | 2 | 4 | Under this project, NCRB is looking to develop the analytics on un-structured data of FIR (First Information Report) captured & stored in 9 different languages such as English, Hindi, Marathi among others.

The required solution must automatically classify the FIRs under different Crime category/ sub-category/ sub-sub-category based on the information stored in free-flow text of FIRs with the help of Machine Learning and Natural Language Processing (NLP) methods & technologies. | 1. As understood, the selected bidder will carry out analytics for two languages of English & Hindi of one lakh data each.

2. As we understand the proposed Data Analytics will primarily focus on the free flow text of FIR Field 12 - First Information contents). Apart from this, if any other field of FIR will be included for Data Analytics. | NLP will be on field 12 of FIR. Selected vendor Is free to use other fields of FIR also. |
| 102 | | 3 | 5 | Project Duration for project roll out | 1. Only three months (90 days) is allocated for doing the entire roll out of the project. However, no clarity on user acceptance process to complete the 2 Lakhs data. If sample data will be verified, the % (percentage) of sample data to be verified may be shared. | Please refer section B of point 11 in RFP |

| 103 | | 4.3 | 5 | As the NLP based classification will be built on the English keywords. For transliteration of text of used languages in FIR to English, NCRB will provide the CDAC supplied APIs (Application Programming Interfaces) for language translations. | 1. It is understood that the transliterated of local language in English text shall be provided using the CDAC provided API of existing CCTNS data. The transliterated text will be shared with the bidder to do the ML/NLP based search of free flow data of FIR field 12 data. Request confirmation. 2. As per RFP 100,000 Hindi comments would be given and API link of CDAC will be used to convert those sentences to English. However, if the quality of the conversion is not good training Data would be below par inadvertently affecting the Model accuracy, would this be taken into consideration? 3. Quality of transliterated data is critical for Data Analytics outcome. Client will ensure the data quality before sharing with the selected bidder for Data Analytics to meet the SLA criteria defined in Para ii (sub para 3) of page 8. | 1. Selected vendor will be given CDAC Transliteration Engine which has to be used by it. NCRB will not provide transliterated text. 2. Sample data will be given to bidders as per serial no. 21. Regarding Quality of data refer serial no. 22 3. As per serial no. 22 |

104		4.7	5	NCRB will provide the testing and deployment infrastructure in their premises in Delhi/NCR.	1. It is understood that NCRB will provide the User Acceptance Testing environment with necessary IT infrastructure & necessary accessories. Bidder will deploy the Data Analytics tool to do the ML/NLP.	As per serial no. 82 and 87
105		4.9	5	Bidder will be provided necessary H/W with required software/database and storage where bidder will be able to store the NLP/ML based analytics data. Database or any other S/W tools will be part of the bid.	1. It is understood that Client will provide IT H/W, S/W, Database, storage along with necessary infrastructure. Bidder will provide the Data Analytics tools and related dependency software for successful execution of work.	As per serial no. 82

| 106 | | 5.4 | 6 | Bidder shall have minimum two resources who have experience of Police System and ML/AI/NLP. | Since the project would require both AI/ ML experience as well as the domain experience. The firm has extensive experience in both the areas and would want to deploy both types of resources to leverage the expertise to the optimum. Requirement of both the domain and Technical skills in the same resource not only would be restrictive but would also impair our ability to put the best resources on the project. Hence, it is requested that the condition 4 of Pre-Qualification may be replaced by following –

1. Bidder shall have minimum two resources who have experience on core application system of CCTNS with 05 years' experience and.

2. Bidder shall have minimum two resources who have experience of ML/AI/NLP with 05 years of experience | No Change |

107		4.4	5	NCRB will provide the keywords-based decision matrix to tag the existing FIR records to different categories/ sub- categories / sub-sub-categories which will be used for training of NLP based ML System.	1. While creating training data, labelling would have to be done. In order to get these category labels, along with the keywords, will bidder be provided with the mapping rules or instructions to help us to link the corresponding FIC (First Information Content) of the FIR?	As per serial no. 2
108		4.11	5	NCRB will use its existing resources to development of reports/graphs on the processed data from this system.	1. It is our understanding that no module needs to be developed for extracting & structuring PDF Data (FIC contents), structured Data in the form of tables would be provided. The same is to be confirmed.	yes

| 109 | | 5.4 | 6 | Resume of at least 2 resources being deployed for the work with below experience and capabilities

§ Educational Qualification: PhD, M.Tech. or B.Tech/BE in Computers or Mathematical computing (Maximum 5 marks)

§ Minimum 5 Years of Experience (Maximum 5 marks)

§ Technology experience of Machine Learning, Natural Language Processing and Artificial Intelligence (Maximum 10 marks)

§ Domain experience of applying relevant technologies in Crime Analytics (Maximum 10 marks) | Requirement of relevant technology experience in Crime Analytics projects would exclude those valuable resources who have immense experience in both AI/ML and Crime Data but in different projects. Hence it is requested that Technical Score S.No.4 may be replaced with following: Resume of at least 2 resources being deployed for the work with below experience and capabilities

§ Educational Qualification: PhD, M.Tech. or B.Tech/BE in Computers or Mathematical computing (Maximum 5 marks)

§ Minimum 5 Years of Experience (Maximum 5 marks)

§ Technology experience of Machine Learning, Natural Language Processing and Artificial Intelligence (Maximum 10 marks)

§ Domain experience of Crime Data. (Maximum 10 marks) | No Change |

110		5	6	Total Score =100	1. Component wise scores do not add up to 100.	Maximum Technical score is 110 instead of 100 as mentioned at section - 6 of RFP. Technical evaluation will be done out of 110 marks. Later during calculation the mark obtained out of 110 by a bidder will be normalized on the scale of 100 and will be used in section 9 (b) of RFP. Additionally, a vendor has to obtain 60% as the technical score for qualifying for the opening of its commercial bid.
111		9	7	a) As this is latest and deep technology work, bid will be evaluated on QCBS (Qualify and Cost Based Selection) considering 60% weightage for technical and 40% weightage for commercial.	1. Keeping the importance of this project in mind, the quality is important and hence more weight of technical score should be given. Please revise this to QCBS 80: 20.	No Change
112		5	6	Additional Suggestion	1. Following clause may be added to Pre-Bid Qualification Condition - Bidder must have at least 100 full time consultants on its payroll.	No Change

113		A) 17 Indemnification/ Page 31		• Subject to clause mentioned below, Successful bidder (the "Indemnifying Party") undertakes to indemnify CLIENT (the "Indemnified Party") from and against all losses on account of bodily injury, violation of intellectual property rights of third party, computer viruses, death or damage to tangible personal property arising in favour of any person, corporation or other entity (including the Indemnified Party) attributable to the Indemnifying Party's performance or non-performance under this Agreement or the SLA to the extent of the Indemnifying Party's comparative fault in causing such losses.	1. The said clause is requested to be replaced by following. • Subject to clause mentioned below, Successful bidder (the "Indemnifying Party") undertakes to indemnify CLIENT (the "Indemnified Party") from and against all losses on account of bodily injury, violation of intellectual property rights of third party, death or damage to tangible property arising in favour of any person, corporation or other entity (including the Indemnified Party) attributable to the Indemnifying Party's performance or non-performance under this Agreement or the SLA to the extent of the Indemnifying Party's comparative fault in causing such losses.	No Change
114			12	Additional Suggestion	1. A format for Power of Attorney may be provided.	Any standard text of Power of Attorney as per India context may be used.
12. SISL Infotech						
115					It may not be possible to achieve 70% accuracy since no Libraries / API (AWS/ GCP) can read free hand text and match the accuracy up to that level.	No Comments

116					What are the formats and quality of data (ex: Pdf/Scanned Image)?	Text format stored in Relational Database.
117					What all programming languages (Python, Java, etc...) does have the ability to call the functions of CDAC Translation Api?	APIs can be called from Industry standard languages. The details will be shared with the selected vendor.
118					What different kind of analysis does NCRB required from NLP Model (ex: reports, dashboard, aggregation, etc...)	Categorization of FIRs
119					What database do the CCTNS system use and their dimensions?	MSSQL/MySQL/Sybase.
120					Reports/graphs will be developed by NCRB or bidder?	NCRB
121					What is the compatibility/dimensions of NCRB's intranet/web-based environment?	NCRB has industry standard internet/intranet/web based environment system. However outside access is very limited as per security protocols of NCRB.
13. Daffodil Software Pvt. Ltd.						
122		Genera l	NA		Please elaborate on type of analytics that will be required from the solution	FIR Categorization

123		Section 2 of RFP Doc	4		In what format are these report captured and maintained? Will it require any optical character recognition and or intelligent character recognition including handwriting?	Free flow text stored in the Relational database. NLP has to be applied to a columns containing FIR text. No OCR required
124		Section 2 of RFP Doc	4		Do you have the list of categories and subcategories handy? Can you please provide an idea of high level numbers..	These will be shared with the selected vendor. However list can grow as per the need of NCRB and provided solution must be flexible enough to accommodate new categories/sub/sub-sub categories.
125		Section 2 of RFP Doc	5		Can you please elaborate "comprehend free-flow text data such as field 12 of CCTNS' FIR data?"	FIR field# 12 contains the free flow text in multiple languages. Sometimes mix language like Hindi+English is also used.
126		Section 4 of RFP Doc	6		Please provide list of high level fields that are captured in a typical FIR.	As per serial no. 97
127		Section 4.3 of RFP Doc	6		What is the accuracy level of your transliteration APIs?	As per serial no. 30 & 103
128		Section 4.4 of RFP Doc	6		Can you please provide some sample keyword based decision matrices that the system would require? Please provide the stats on possible number and permutation and combination of scenarios to start with. What	As per serial no. 2

						level of nesting would be required?	
129			Section 4.11 of RFP Doc	6		Should we assume that the reporting module in not included in the current scope and We will just push the data in a given format (REST/SOAP) to NCRB platform?	Yes, however how the data will be be pushed/stored in NCRB system will be discussed with the selected vendor.
14. Infosys							
130			Section 2 - Page 5		FIR Related Details	Can some sample FIR documents be provided with appropriate redaction of sensitive data?	As per serial no. 97
131			Section 2 - Page 5		FIR Related Details	What percentage of FIRs would be handwritten?	Very few and they are not in the scope of this project.
132			Section 4: #4.3 - Page 6		CDAC Capability	It is our understanding that CDAC tool will transliterate Hindi words into English characters while retaining the same pronunciation. For e.g. नमस्ते as Namaste. Pl. confirm if the understanding is correct.	Yes

133			Sectio n 4: #4.3 - Page 6	CDAC Capability	In addition to transliteration, is CDAC also capable of translating the non English Languages into English?	No
134			Sectio n 4: #4.4 - Page 6	Tagging using keywords	Will the keyword based decision matrix also include support for transliterated keywords such as Chori, Qatl etc.?	As per serial no. 2 which will include words like Chori and Qatl
135			Sectio n 4: #4.5 - Page 6	Analytics	Is the analytics System mandatory requirement as a part of this RFP?	NLP and ML based analytics for processing FIR data only
136			Sectio n 11 - Page 9	Categorization Accuracy	Are the accuracy threshold open to revision?	No
137			Sectio n 11 - Note B - Page 10	Categorization Accuracy	Who would be doing accuracy check – NCRB or Bidder?	NCRB
138			Page 5 : 4.6	Scope of Project	Bidder should use open source technologies for development of the system. Please clarify	As per serial no. 1
139				Scope of Project	Any company can participate or CMM5 level company with certain Annual revenue category?	Any company can participate as per prequalification criteria mentioned in the RFP
140					Does this project require support post this implementation?	Warranty of 5 Years

141			Pre-Qualification Criteria	Bidder should have executed 5 projects in the area of Machine Learning or Artificial Intelligence or Natural Language Processing. Bidder to provide copy of orders or completion certificate	Please accept work experience certificate from bidder Authorized signatory/ CFO / Company Secretary. For global project citations, due to NDA reasons, it will be difficult to provide work order and completion certificate.	Bidders can submit evaluation committee will scrutinize
142			Pre-Qualification Criteria	Bidder shall have minimum two resources who have experience of Police System and ML/AI/NLP.	Please remove the experience specific to Police system to allow larger participation.	No Change
143			Pre-Qualification Criteria	Certifications	Please include the Bidder should have valid CMMI Level 5 and a valid ISO 9001: 2015 or ISO 27001:2013 as on the date of submission of the bid.	No Change
144			Pre-Qualification Criteria	Financials	Please include the Bidder should have Annual Turnover of minimum INR 100 Crores. During the three preceding financial years (FY) : FY 2016-17, 2017-18, 2018-19	No Change
145			Standard Terms and Conditions		Please include the bidder total liability under the Contract shall be limited to 100% of the Contract Value.	No Change

15. Sterling Varnishes (Bangalore) P. Ltd.						
146					1. Would you need some level based Access to the final application? Or Complete access?	Complete
147					2. The final Roll out would be used by how many people simultaneously? Approx.	Many users but the application must not restrict it usage.
16. Ad Bureau Advertising Pvt Ltd.						
148					1. How many categories would be present (approx.) in FIR data.	As per serial no. 124
149					2. The input dynamic queries would be in English or Hindi as well.	English & Hindi