HP eDiscovery Best Practices Meaning Based Coding Implementation Process

Contents

Introduction 2
Predictive Coding-Based Review 2
Control Set 2
Seed Set 3
Random Sampling 3
Judgmental Sampling 4
Control and Seed Set Validation 6
Use Case 1 – Iterative Training 6
Use Case 2 – Accelerated Relevance Review 7
Linear Review Optimizations 8
Bypass Tier 1 Review – Relevant 8
Tier 1 Review – Non-Relevant QC 9
Pre-Production Relevance QC 9
Pre-Production Privilege QC 9
Pre-Review Culling 9
Basic Suggested Coding Display 10
Viewing Suggested Coding 10
Predictive Coding Metrics 11
Definitions 11
Metrics and Reporting 12
Summary 13

Introduction

HP Autonomy's business white paper, *Meaning Based Coding in HP Autonomy eDiscovery*, provides a concise summary of the proprietary Technology Assisted Review (TAR) solution in HP eDiscovery, Meaning Based Coding (MBC). It explains how MBC achieves the goals of predictive coding – time and cost efficiencies and higher-quality review coding – in a measurable and defensible manner.

This document is for existing clients who are interested in learning in greater detail how predictive coding can be leveraged on their existing or future projects. Unlike other vendor solutions that lock clients into a rigid methodology and review workflow, one of the advantages of MBC is its flexibility in supporting multiple predictive coding use cases. MBC can be leveraged to reduce the number of documents requiring review or to optimize and provide quality control options on a linear review.

From a process standpoint, HP eDiscovery provides flexibility in supporting "real world" use cases. Most eDiscovery vendors address "perfect world" scenarios in their predictive coding marketing papers, but real eDiscovery projects are characterized by complexity, changing requirements, and opposing party demands. Can different review approaches be used depending on the data source? How should document families be handled? What is the best workflow for documents without text? How does the collection and processing of ESI on a rolling basis impact predictive coding?

Because of project-specific requirements and challenges, HP eDiscovery's professional services team takes a consultative approach with hosting clients in implementing and revising MBC-based solutions. Technology, Process, Services, and Support – your success depends on a vendor that has all the bases covered.

Predictive Coding-Based Review

MBC's support of two widely accepted use cases for predictive coding-based document review is detailed below. The common goal of these use cases is to reduce the number of documents manually reviewed, while still identifying all relevant documents within an acceptable, transparent, and approved margin of error.

The HP eDiscovery implementation of these use cases is consistent with the EDRM's Computer Assisted Review Reference Model. Both use cases require the creation of a Control Set of documents that are reviewed to establish a statistical baseline, against which MBC's predictions are compared, as well as the creation and review of a Seed Set to initiate the process. It is important to thoroughly understand the role of, and implementation methods for, the Control Set and Seed Set.

Control Set

The Control Set is drawn from a statistical random sample of un-reviewed documents with extracted text. The Control Set is manually reviewed and tagged by expert reviewers to establish a statistical baseline, e.g. percentage of relevant documents, which is used to extrapolate how many relevant documents are expected in the document population.

If additional data sources for existing or new custodians are collected and processed after the initial Control Set is created, the Control Set is refreshed with a new random sample drawn from the expanded document population. The refresh is necessary since the new data sources may contain net new relevant concepts.

Seed Set

MBC processing on a selected work product tag is based on two parameters specified by the client. The first parameter is the Coding Criteria, otherwise called the Training Set. The second parameter is the Target Universe. MBC makes coding predictions for the documents in the Target Universe based on the manual coding of the documents in the Training Set. The composition of the Training Set and the Target Universe is revised during the course of review depending on which predictive coding use case is implemented, but the initial Training Set is often referred to as the Seed Set.

The Seed Set in turn is usually composed of a Random Sample and a Judgmental Sample, which is a set of documents, known as "exemplars", that are either highly relevant and/or highly non-relevant. Such documents are intended to "jump start" the MBC process in making high-quality predictions as soon as possible.

Like the requirement to refresh the Control Set if new data sources enter the review population, the Seed Set should be augmented with additional judgmental sampling drawn from the new data sources. The net new concepts identified in the new data sources are not only necessary to help make predictions for documents within these data sources, but they will help refine the predictions on previously processed data sources.

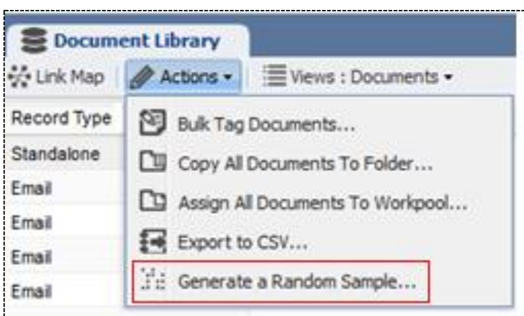
Note: Though some clients elect for simplicity to use the same set of documents for the Control Set and Seed Set, some predictive coding consultants and literature in the market recommend distinct sets. This decision can be left up to the client.

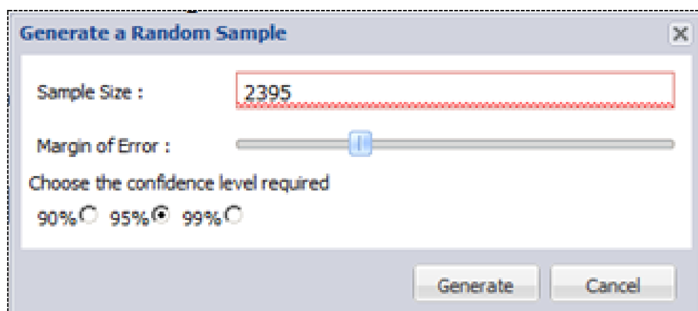
Random Sampling

An industry-standard and defensible technique to identify documents for the Control Set and Seed Set is to draw a statistical random sample from the document population to be reviewed.

As illustrated below, the name of the feature in HP eDiscovery is "Generate a Random Sample," and it is available on the Document List "Actions" menu to authorized users. The user can specify a confidence level and margin of error (which auto-calculates the displayed sample size) or directly enter a sample size to use.

In addition to using this feature to populate the Control Set and Seed Set, this feature is also used on many of the use cases to draw random samples for predictive coding validation.





Whereas a definitive, best practice confidence level for predictive coding random sampling has yet to be established, a value of either 95% or 99% (within a margin of error of 2%) is generally accepted. Some clients prefer to round up an auto-calculated size, e.g. specify 2,400 as the sample size at the 95% confidence level. The legal team must choose the confidence level and margin of error to use in all random sampling, usually with the agreement of opposing counsel and/or the court.

As listed below, at a confidence level of 99%, a document population of 1M documents requires a sample set of only 4,143 documents. It may seem counterintuitive that if the population grows 50% to 1.5M documents, the sample set only needs to increase by 6 documents, but such is the nature of statistical random sampling.

Confidence Level	Margin of Error	# of Documents					
		100K	250K	500K	750K	1M	1.5M
95%	2%	2,345	2,378	2,390	2,393	2,395	2,397
99%	2%	3,994	4,092	4,126	4,137	4,143	4,149

The review team needs to decide whether to include family members of the random sample when assigning documents to the random sample review team. Including family members increases the number of documents to review, but the client might want the reviewers to have some context to guide the tagging decision, especially in the case of attachments.

The review team must also take care in specifying the set of documents to be sampled. Ideally, all documents within the scope of the review would be processed and de-duped before the random sample is taken. On a typical eDiscovery project however, data is collected, processed, and reviewed on a rolling basis.

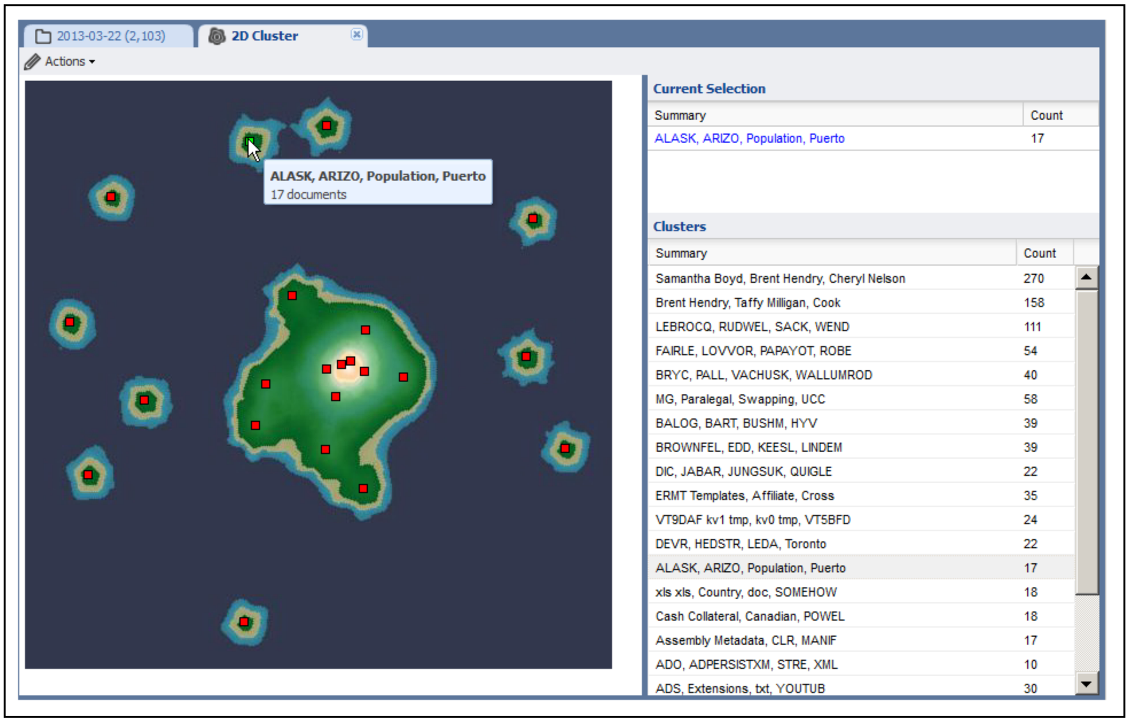
Consider an employee slip and fall litigation with five custodians: Operations VP (email), Supervisor (email), Employee (email), HR Policy Documents (file server), and Witness (email). If the Operations VP and Supervisor emails were collected and processed first, the random sample will be representative of the relevant concepts across these two custodians. But what if there are additional relevant concepts that are only present in the other three custodians' documents? Therefore, on projects where MBC processing starts before all custodial data sources are collected and processed, the Control Set and Seed Set are refreshed periodically as new data sources are processed and released for review.

Judgmental Sampling

Judgmental Sampling aims to identify "exemplars" – highly relevant and highly non-relevant documents – to include in the Seed Set. HP eDiscovery provides a variety of judgmental sampling techniques.

<u>Clustering</u>

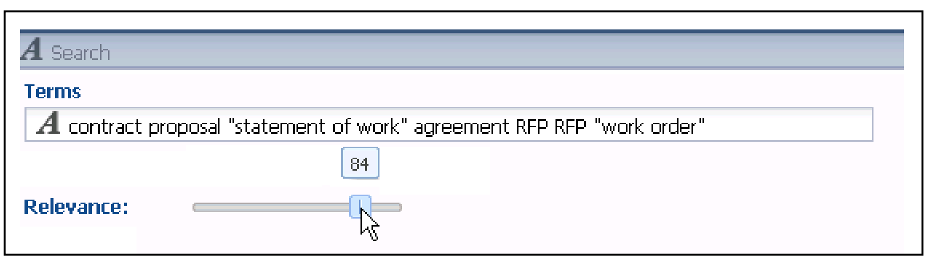
The clustering feature is used to identify conceptually similar subsets of documents within a specified document population. These subsets or "nodes" can be reviewed and tagged for inclusion in the Seed Set. Clustering may not only expose highly relevant documents to include in the judgmental sample, but highly non-relevant documents. An Enron data set cluster map is illustrated below.



<u>Conceptual Search</u>

Most clients develop a “term list” to identify potentially relevant documents, and this list of keywords and/or full text search expressions is often agreed upon with opposing counsel for culling the document population for review.

If you search on all or a subset of these terms in the HP eDiscovery search interface while setting a minimum relevance rank for the search results, some of the search results with the highest relevance ranks might be exemplars to add to the Seed Set.



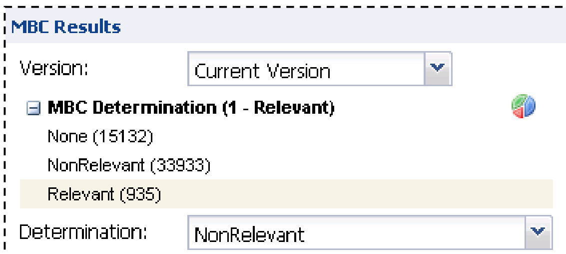
The review team should also consider other advanced searches and HP eDiscovery analytical techniques, e.g. Automatic Query Guidance, Visual Email Analytics, to identify possible exemplars.

While judgmental sampling, remember that non-relevant exemplars can be as useful in a Seed Set as relevant exemplars. For example, on a project where a search on the term "credit swap" returns potentially relevant documents, an expert reviewer could review the search results and include those highly relevant in the Seed Set. But if "Credit Swap" was also the name of the company softball team, the reviewer might want to identify highly non-relevant documents discussing the softball team and include them in the Seed Set.

Control and Seed Set Validation

Expert reviewers should review both the Control Set and Seed Set, but even expert reviewers can incorrectly code similar documents. Since it's important to have these sets as accurate as possible, MBC can be used where the Training Set and the Target Universe are the same set (self-evaluation method).

Control and Seed Set reviewers can compare the MBC-predicted coding with their manual coding. They can filter on the MBC coding fields to display just the discrepancies. For example, the settings below will filter a document list to display all documents coded non-relevant by a reviewer, but predicted relevant by MBC. The reviewers can then adjust their previous coding if necessary.



Use Case 1 – Iterative Training

Iterative Training is the use case most commonly associated with predictive coding based review. This approach aims to "train" the system through an iterative approach until the system is able to predict the subset of the document population that is relevant. The size of this subset is the percentage of documents identified as relevant in the Control Set, subject to the margin of error.

The first time, or "iteration", the MBC process is run, the Target Universe is a large random sample, e.g. no more than 200K documents, drawn from the review population. (Note: the larger the target universe, the longer it takes for MBC to complete iterative processing.) When creating the Target Universe Set, care is taken to (a) exclude documents without text and (b) exclude documents in the Control Set.

After the MBC process has predicted coding for the documents in the initial Target Universe, expert reviewers manually review and code them to assess the accuracy of the predictions. After each training round, metrics including Precision, Recall, and F1-Score are updated to provide insight as to how quickly the system is learning, i.e. providing ever improving predictions.

The system cannot "learn" however unless it's been informed of its past mistakes. Therefore, the documents that were incorrectly predicted or left unclassified (predicted coding value of 'None') by MBC in the previous iteration are added to the Training Set for the next iteration.

A new Target Universe Set is created and the process is repeated until (a) the percentage of documents predicted relevant in the iteration matches the percentage of documents predicted relevant in the Control Set (with the chosen margin of error), and (b) the percentage of documents unclassified by MBC is determined acceptably low.

When the system is sufficiently "trained", i.e. the Training Set contains an ideal set of documents for making predictions, the MBC process is run with the Target Universe set to the remaining, un-reviewed document population.

Next, the predictive coding results need to be validated using the same assessment methodology used in the iterative runs. The Target Universe is checked to see if (a) the percentage of documents predicted relevant matches the percentage of documents predicted relevant in the Control Set, and
(b) the percentage of documents unclassified by MBC is determined acceptably low.

Statistical random sampling is also used for validation. A random sample of the Target Universe is drawn and reviewed. A specific metric critical to defensibility, Elusion, is calculated. Elusion is the percentage of documents in the random sample of all documents predicted non-relevant that the reviewer determines are actually relevant. The legal team needs to determine the acceptable cutoff percentage.

Advanced validation techniques to consider include systematic or interval sampling. A specific approach is to take multiple random samples of documents within different prediction confidence score ranges, e.g. <= 25% | > 25% and <= 50% | > 50% and <=75% | > 75%. If, for example, relevant documents outside the margin of error are found in the sets of documents with the lowest confidence scores, additional training iterations are required with these documents added to the Training Set. Another technique to consider is drawing random samples for validation on a per custodian basis.

Once the training iterations and validations are complete, the document population can move forward in the workflow as outlined below. The work product tag needs to be bulk-coded with the coding value predicted by MBC; the tag can always be revised in a subsequent review tier if necessary.

Standalone documents or full document families that MBC did not classify (predicted coding value of 'None') enter Tier 1 Review, as clients usually elect to have lower-cost Tier 1 reviewers make the initial call on these documents.

Document families that contain at least one document predicted relevant enter Tier 2 Review or Privilege Review before production. The client may optionally specify a prediction confidence score "cutoff", where at least one of the relevant document's confidence score needs to be higher than the cutoff for the document family to enter Tier 2 Review or Privilege Review. The document families that contain relevant documents whose confidence scores are all below the cutoff enter Tier 1 Review.

After the unclassified and relevant document families described above enter their appropriate review workflows, the remaining documents in the un-reviewed population are document families predicted non-relevant by MBC. The remaining non-relevant families do not require further review (hence the project cost and time savings of predictive coding-based review), though the client may optionally specify a prediction confidence score cutoff, below which document families enter Tier 1 Review to confirm the non-relevant prediction.

Use Case 2 – Accelerated Relevance Review

Accelerated Relevance Review is a more "conservative" predictive coding approach than Iterative Training in that a Tier 1 Review of potentially relevant documents is still required. The Tier 1 Review however is accelerated by leveraging MBC's predicted coding.

In this use case, the Target Universe is always the remaining, un-reviewed document population, and the initial Training Set is the Seed Set. After the first MBC iteration is run, document families that contain documents that MBC predicts are relevant enter Tier 1 Review.

The client may optionally specify a prediction confidence cutoff score to reduce the set of documents entering Tier 1 Review, and therefore ensure more frequent iterations. There are dual benefits in this approach. First, Tier 1 reviewers will review faster because it takes less time to confirm the suggested coding of a document with a higher versus a lower confidence score. Second, whenever reviewers code unclassified documents as well as identify false-positives (documents MBC suggested were relevant but are actually non-relevant) and false-negatives (documents MBC suggested were non-relevant but are actually relevant), these documents are added to the Training Set to enable MBC to make better predictions on the next iteration.

A recommended option to achieve further acceleration is to specify a distinct prediction confidence score cutoff, where documents predicted relevant above the cutoff, and their family members, skip Tier 1 Review and enter either Tier 2 Review or Privilege Review. This option assumes that no other coding by Tier 1 reviewers is required, e.g. key document and/or issue tagging.

If review is driven by likely relevant documents entering review first, then as the remaining documents left to review nears exhaustion, the percentage of documents that are confirmed as non-relevant will increase. The process is repeated until (a) the overall percentage of documents predicted relevant in document population matches the percentage of documents predicted relevant in the Control Set (+/- 2%), and (b) the percentage of documents unclassified by MBC (suggested coding value of 'None') is determined acceptably low. Once this milestone is reached, further Tier 1 Review can be discontinued.

For example, consider a document population of 1M documents from which the original MBC Control Set is created with a 95 percent confidence level and a 2 percent margin of error. If 20% of the Control Set is tagged relevant, then within the 2 percent margin of error it can be extrapolated that between 180K and 210K documents in the entire population are relevant. Therefore, once between 180K and 210K documents are tagged relevant, the client can discontinue Tier 1 Review. Remember, this approach does not guarantee that there are no relevant documents left in the document population, only that the likelihood is within an acceptable margin of error.

The same random sample validations against the remaining, un-reviewed document population and the Elusion metric calculation documented in the Iterative Training use case above need to be performed in this use case. The review team needs to verify and defend the decision to not manually review the set of documents that MBC has predicted as non-relevant.

Linear Review Optimizations

Implementing a predictive coding review workflow requires a "leap of faith" that many companies and law firms are not yet ready to take, though such reticence is declining as predictive coding continues to gain acceptance. Research papers by independent information retrieval experts conclude that predictive coding is more accurate than human review alone, but old attitudes and existing review methodologies are entrenched.

Therefore, sometimes a company, law firm, or document review vendor needs to "ease into" using predictive coding techniques on a linear review to gain confidence and fully appreciate the efficacy of predictive coding in general. MBC is engineered to support a variety of linear review optimizations and QC use cases.

In these use cases, a Seed Set is required as described above. The updates to the Training Set and the Target Universe, however, vary depending on the specific use case selected for implementation.

Bypass Tier 1 Review – Relevant

This use case is based on the Determination tag. Document families that MBC determines contain at least one relevant document bypass or skip Tier 1 Review. However, it is not HP Autonomy's recommendation to have documents that bypass Tier 1 Review subject to immediate production. A Privilege Review and/or Tier 2 Review to confirm the suggested relevant designation is recommended.

The HP Autonomy services team will consult with the client on the precise details of implementing this use case. For example, the client will be advised on the decision to have all document families for which a relevant determination was suggested by MBC bypass Tier 1 Review, or only those families where at least one document achieved a high confidence score.

All documents that MBC suggests are non-relevant still pass through the linear Tier 1 Review.

Tier 1 Review – Non-Relevant QC

The standard approach to performing QC on document families exiting a standard Tier 1 Review with a non-relevant designation (all family members) is to perform a periodic, random sample QC review. MBC provides an additional method however.

On a periodic basis or once at the end of project review, all non-relevant document families that exited Tier 1 Review are checked using MBC. Documents containing discrepancies between the manual coding and MBC's predicted coding are directed to a specialty reconciliation review, or simply redirected into the standard Tier 2 Review for definitive tagging. Any documents within this set that require coding changes, e.g. false-negatives identified and corrected, are ideal documents to add to the Training Set to increase the accuracy of future iterations.

Pre-Production Relevance QC

A typical goal of document review is to produce relevant documents to the opposing party per a legal or regulatory obligation. Clients usually perform a two tier review to ensure all necessary relevant documents are produced. But there will often be non-relevant documents included in production because of coding mistakes made in the Tier 1 review not caught in subsequent review tiers.

Similar to how MBC can be used to perform post-Tier 1 Review on document families tagged non-relevant, a pre-production QC workflow can be implemented to identify any non-relevant documents that are about to be inadvertently produced. On a periodic basis, all or a sample of relevant document families exiting Tier 2 Review and Privilege review are compared against MBC processing results.

Documents containing discrepancies are directed to a specialty reconciliation review for coding validation or retagging if necessary. All documents with discrepancies are ideal documents to add to the Training Set.

Pre-Production Privilege QC

This use case is based on whatever Privilege tag has been implemented. Ideally, all privilege documents are identified during Tier 2 Review and/or Privilege Review. MBC is used to identify documents incorrectly tagged not-privileged during a manual review. All documents containing discrepancies are directed to a specialty reconciliation review for coding validation or coding changes, if necessary. All documents with discrepancies are ideal documents to add to the Training Set.

Pre-Review Culling

The MBC use cases detailed above are typically implemented against the post-culled document corpus. On many eDiscovery projects, pre-review culling is performed using a list of keyword terms and/or boolean search expressions agreed upon between the parties. However, if it is of strategic benefit to the client, MBC could be proposed and implemented as a more accurate culling method. A minimum confidence threshold could be established to determine which document families are withheld from review. A periodic, random sample QC review of withheld documents is recommended.

Basic Suggested Coding Display

In this simplest of use cases, the standard, linear review workflow remains the same. The only change is that MBC is used in parallel to generate suggested coding values that the review team can optionally consider. If you do not want to bias the Tier 1 reviewers by displaying the suggested coding values, the display of the values could be limited to select reviewers, e.g. Tier 2 and/or Privilege reviewers.

In support of this use case and as required by review managers, the section below describes how authorized users can display the MBC predictive coding results.

Viewing Suggested Coding

The MBC process is applied on one or more work product tags including the Determination tag (the HP eDiscovery standard tag used for the relevance call), the Key Document tag, the Privileged Status tag, and the multi-value Issues tag, though Determination is the most frequently selected tag.

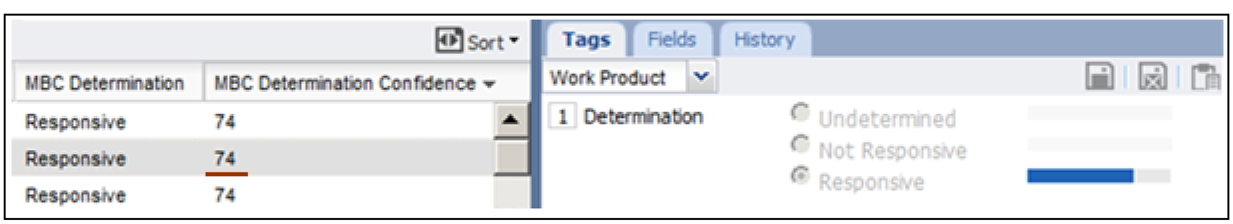
For a work product tag selected for MBC processing, the suggested coding values and confidence scores are stored in two system-generated fields. The suggested coding is stored in a field named "MBC" + <tag name> and the confidence score is stored in a field named "MBC" + <tag name> + "Confidence".

HP eDiscovery makes it easy for authorized users to display and leverage these fields using one or both of the methods illustrated below.

<u>Document List</u> – one or both MBC result fields can be selected by a user for display in a document list

Record Type	Determination	MBC Determination	MBC Determination Confidence
Email	Responsive	Responsive	100
Attachment	Responsive	Responsive	100
Attachment	Responsive	Responsive	100
Attachment	Responsive	Responsive	100
Attachment	Not Responsive	Not Responsive	81
Attachment	Responsive	Responsive	100
Attachment	Not Responsive	Not Responsive	81

<u>Predictive Coding Bars</u> – an optional confidence score bar can be displayed in the coding panel



Meaning Based Coding Implementation Process

The length of the colored bar, relative to the fixed bar length, illustrates the confidence score.

Predictive Coding Metrics

Measuring the MBC process is important both for workflow management as well as to document the process for defensibility. There are a number of predictive coding-specific performance metrics that are gaining acceptance in the industry; the reader is encouraged to reference the EDRM website to learn more.

Definitions

The metrics used to measure the effectiveness of a TAR solution are based on the following four key metrics generated after every MBC iteration.

True Positive (TP)	**MBC correctly predicted as positive, e.g. relevant (if that's the selected tag)**
False Positive (FP)	MBC incorrectly predicted as positive
True Negative (TN)	MBC correctly predicted as negative, e.g. non-relevant
False Negative (FN)	MBC incorrectly predicted as negative

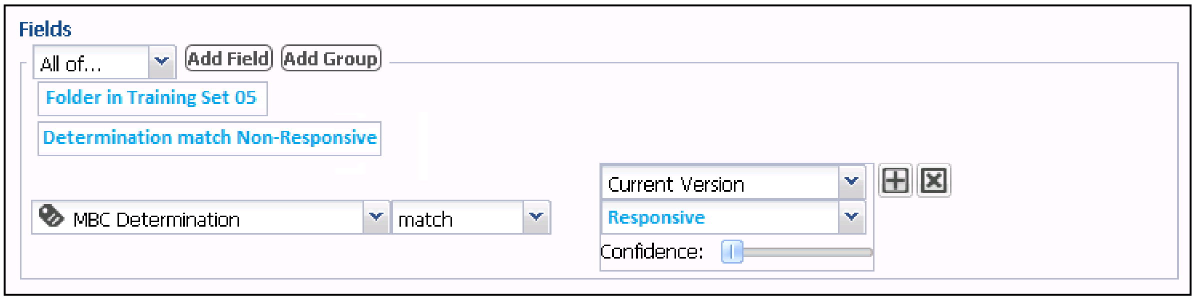
The entire set of performance measures based on these four metrics is defined in the table below, though Precision, Recall, F1-Score, and Elusion are the most commonly referenced.

Precision	% of documents predicted positive that are indeed positive. Also referred to as Positive Prediction Value (PPV).
Recall	% of positive documents correctly predicted as positive = 100% - FNR
False Negative Rate (FNR)	% of positive documents incorrectly predicted as negative
F1-Score	Harmonic mean weighing Recall and Precision
F2-Score	Similar to F1 but with greater weight given to Recall: $F_B = (1 + B^2) * ((Precision * Recall) / ((B^2 * Precision) + Recall))$
Elusion	% of documents predicted as negative that are actually positive. Elusion is best measured by taking a random sample of the set of documents predicted as negative. If the Elusion value is sufficiently low, iterations can be stopped.
Negative Prediction Value (NPV)	% of documents predicted negative that are indeed negative = 100% - Elusion
Error	% of documents inaccurately predicted
Accuracy	% of documents accurately predicted or 100% - Error
Prevalence	% of positive documents. Also referred to as Richness or Yield.
False Positive Rate (FPR)	% of negative documents incorrectly predicted as positive
True Negative Rate (TNR)	% of negative documents correctly predicted as negative 100% - FPR

Table citation: The Grossman-Cormack Glossary of Technology-Assisted Review. (2013). www.edrm.net.

Metrics and Reporting

In HP eDiscovery, it is easy to calculate the values of the four base metrics (TP, FP, TN, and FN). For example, if you need the False Positive (FP) value for a Training Set that has been reviewed, you need to identify all documents within that set that MBC predicted were *relevant* that the validation reviewer coded *non-relevant*. The number of documents returned by the following HP eDiscovery search is therefore the False Positive (FP) value.



The following set of metrics is used to track predictive coding performance:

# of documents in the Control Set
# of documents in this set coded True (Relevant)
# of documents in this set coded False (Non-Relevant)

# of documents in the Training Set
# of documents in this set coded True (Relevant)
# of documents in this set coded False (Non-Relevant)

# of documents in the Target Universe
# of True Positives (TP) – Truly Relevant and Predicted Relevant
# of False Positives (FP) – Truly Non-Relevant but Predicted Relevant
# of True Negatives (TN) – Truly Non-Relevant and Predicted Non-Relevant
# of False Negatives (FN) – Truly Relevant but Predicted Relevant

The performance measures tracked are illustrated below.

<u>Control Set</u> – the basic Control Set metrics are tracked across iterations in case the set is revised due to changes in the document population over time

		Control Set		
Iteration		**# Docs**	**Tagged True**	**Tagged False**
1		4,000	388	3,612
			9.7%	90.3%
2				

Meaning Based Coding Implementation Process

Training Set – the Training Set is the MBC coding criteria that is specified every time to the MBC process is run; the Training Set increases after each iteration as new exemplars are added

		Training Set		
Iteration		# Docs	Tagged True	Tagged False
1		8,193	544	7,649
			6.6%	93.4%
2				

Confusion Matrix – industry-standard table of TP, FP, TN, and FN values. The MBC version of this matrix includes metrics for the documents in the Target Universe left unclassified by MBC, i.e. documents for which the Training Set did not provide sufficient conceptual insight for MBC to make a prediction.

	Confusion Matrix / Contingency Table						
Iteration		Predicted Responsive	Predicted Non-Responsive	Not Classified ('None')	% Predicted Responsive	% Predicted Non-Responsive	% Not Classified ('None')
1	**Truly Responsive**	4,500 (TP)	10,000 (FN)		9.0%	20.0%	0.0%
	Truly Non-Responsive	500 (FP)	35,000 (TN)		1.0%	70.0%	0.0%

Performance Metrics – the automatically calculated performance metrics defined in the table below.

	Performance Metrics														
Iteration	# Docs	% True	% False	Precision	Recall	FNR	F1-Score	F2-Score	Elusion	NPV	Error	Accuracy	Prevalence	FPR	TNR
1	50,000	29.0%	71.0%	90.0%	31.0%	69.0%	46.2%	35.7%	22.2%	77.8%	21.0%	79.0%	29.0%	1.4%	98.6%
2															

Summary

HP eDiscovery's Meaning Based Coding (MBC) feature is used to implement industry-standard predictive coding use cases, as well as manual review optimization and QC use cases, which are gaining acceptance among customers and vendors alike.